# Volatile Organic Compounds and Antioxidants in Olive Oil: Their Analysis by Selected Ion Flow Tube Mass Spectrometry

*A thesis submitted in partial fulfilment of the requirements for the degree of*

Doctor of Philosophy in Chemistry

*at the*

University of Canterbury
Christchurch
New Zealand


Brett M. Davis
December 2007

# Abstract

The application of Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) to the analysis of olive oil shows several distinct advantages over more conventional analysis techniques. The two areas described in this thesis examining olive oil quality are the analysis of Volatile Organic Compounds (VOCs) and the assessment of antioxidant activity. VOCs are responsible for the aroma and much of the taste of olive oil, while antioxidants afford some protection from harmful reactions involving radical species inside the body by scavenging radicals when olive oil is ingested.

The VOCs of olive oil are used by sensory panel judges to classify oils by their degree of suitability for human consumption. The major parameters used for this evaluation are the strengths of any defects and the degree of fruitiness. A defect is an indication of an undesired process which has occurred in the oil, while fruitiness is a fragile attribute which denotes a good quality oil and is easily masked by defects. SIFT-MS was used to measure the strengths of the olive oil defects rancid, winey, musty, fusty and muddy. Great potential was demonstrated for all defects except musty and the concentrations of VOCs in olive oil head space were correlated with the peroxide value, a measure of the degree of oil oxidation. A study aimed at correlating the strength of the fruitiness attribute as determined by a sensory panel with the concentrations of VOCs in olive oil head space was unsuccessful.

The SIFT-MS Total Oxyradical Scavenging Capacity (TOSC) assay was used to measure olive oil antioxidants. This assay measures all antioxidants in oil, not only those removed by extraction with a solvent, as it is conducted in an emulsion. SIFT-MS-TOSC assay results were found to correlate well with those of the widely used Folin-Ciocalteu assay and the total concentration of phenolic compounds present in olive oil. Discrepancies between the two assays were most likely due to

hydrophobic antioxidants which are measured by the SIFT-MS-TOSC assay but not the other tests.

# Acknowledgements

Over the course of this research I have interacted with many people, not all of whom I will remember to name here. My sincere thanks to all of these people, you made many occasions over the past few years pleasant when they could easily have been otherwise.

I would like to thank my supervisors, Murray McEwan and Senti Senthilmohan for guidance and valuable advice. Thanks are due to everyone at Syft Technologies, past and present, especially the Science team of those previously named and additionally Greg Francis, Vaughan Langford, Daniel Milligan, Barry Prince and Paul Wilson. Without the equipment, work atmosphere and supplemental employment available at Syft Technologies this research would not have been possible. Thank you to Olives New Zealand, especially Don Beaven, Alastair Bridge, Don Cross, Margaret Edwards and Ian Manson, without whom I would have had no olive oil to analyse. I would also like to thank Helen Clausen, Judy Wilson and Patrik and Eleanor Hulsman for inspiration, advice and enthusiasm. Thank you to all the olive growers and processors who donated olive oil for this project.

This research was funded by a Technology for Industry Fellowship from the New Zealand Foundation for Research Science and Technology. This support was essential and is very much appreciated.

Forewarned is forearmed, and I am grateful to Ben Perston (the Matlab guru) and Michelle Hamilton for forearming me in the later stages of this research. Both also assisted me when wading through the literature of multivariate statistical methods. My sanity and study have also benefited from the presence of David Bones, James Bull and

Sam Edwards. I must also thank the marine group, especially Sonja, Jenny and Annabel for saving me from the HPLC gremlins. Thanks to Wayne MacKay and Nick Oliver in the mechanical workshop for their expertise and tact in dealing with an array of original problems.

Thank you to my family, especially my mother, who (mostly) restrained her constant urge to ask when I was going to finish. Lastly, a big thank you to Kelly for her remarkable tolerance.

# List of Publications

At the time of printing, the following peer-reviewed publications have resulted from this research:

Brett M. Davis, Senti T. Senthilmohan, Paul F. Wilson and Murray J. McEwan, **Major volatile compounds in head-space above olive oil analysed by selected ion flow tube mass spectrometry**, *Rapid Commun. Mass Spectrom.*, 2005, **19**: 2272-2278.

Brett M. Davis and Murray J. McEwan, **Determination of Olive Oil Oxidative Status by Selected Ion Flow Tube Mass Spectrometry**, *J. Agric. Food Chem.*, 2007, **55**: 3334-3338.

LIST OF PUBLICATIONS

# Contents

# List of Abbreviations

| | |
|---|---|
| AAPH | 2,2'-Azobis(2-Amidinopropane) Hydrochloride |
| ABAP | 2,2'-Azobis-Amidinopropane |
| DHS | Dynamic Head Space |
| DTPA | Diethylene Triamine Pentaacetic Acid |
| EDTA | Ethylene Diamine Tetraacetic Acid |
| FFA | Free Fatty Acid |
| FID | Flame Ionisation Detector |
| GAE | Gallic Acid Equivalents |
| GC-MS | Gas Chromatography Mass Spectrometry |
| GLC-MS | Gas-Liquid Chromatography Mass Spectrometry |
| HAT | Hydrogen Atom Transfer |
| HPLC | High Performance Liquid Chromatography |
| $IC_{50}$ | 50 % Inhibition Concentration |
| INST | Instrumental |
| IOOC | International Olive Oil Council |
| KMBA | α-Keto-γ-Methylthiobutanoic Acid |
| LOX | Lipoxygenase |
| MLR | Multiple Linear Regression |
| NIPALS | Nonlinear Iterative Partial Least-Squares |
| ONZ | Olives New Zealand |
| ORAC | Oxygen Radical Absorbance Capacity |
| PCA | Principal Component Analysis |
| PLS | Partial Least-Squares |
| PRESS | Prediction Error Sum of Squares |
| RMSECV | Root Mean Square Error of Cross-Validation |
| R-PE | R-Phycoerythrin |
| SENS | Sensory |
| SET | Single Electron Transfer |
| SHS | Static Head Space |
| SIFT-MS | Selected Ion Flow Tube Mass Spectrometry |
| SIM | Selected Ion Monitoring |

| | |
|---|---|
| SPME | Solid-Phase Microextraction |
| TAG | Triacylglycerol |
| TEAC | Trolox Equivalent Antioxidant Capacity |
| TOSC | Total Oxyradical Scavenging Capacity |
| TRAP | Total Radical-trapping Antioxidant Parameter |
| $VI_{50}$ | 50 % Inhibition Volume |
| VOC | Volatile Organic Compound |

# Introduction

Many techniques exist for the analysis of olive oil. An ideal technique is rapid, reproducible, objective and well understood. Two important areas of olive oil analysis, the evaluation of flavour and antioxidant properties, are not currently investigated using analytical techniques with these characteristics.

Olive oil flavour is important, as it determines the acceptability of the oil for consumers. The flavour is currently evaluated by the use of sensory panels – groups of 8-12 trained judges who compare the flavours of olive oil samples against reference samples to determine the intensities of different sensory attributes. This technique is remarkably reproducible, however it is neither rapid nor well understood.

Antioxidant properties are also important, as olive oil has been shown to contain antioxidant compounds which can lessen the likelihood of cancer or atherosclerosis development in the human body when ingested(*1*). There is currently no standard method to assess the antioxidant properties of any food.

The work presented in this thesis is divided into two parts: one investigating the flavour of olive oil, the other its antioxidant properties. The aim was to develop and evaluate a new experimental technique, Selected Ion Flow Tube Mass Spectrometry (SIFT-MS), for the analysis of olive oil VOCs and antioxidants.

The first two chapters contain background information on olive oil production and composition and a description of the SIFT-MS technique. The data available from SIFT-MS analysis and the information obtained from them are discussed. The distinctive ability of the SIFT-MS technique to provide accurate concentrations of gas-phase analytes in real-time is explained.

The next four chapters describe efforts to correlate the intensities of sensory attributes of olive oil with head space VOC concentrations obtained via SIFT-MS analysis. Chapter three is an introduction outlining efforts that have been made in this area of research.

Chapter four outlines the development of a SIFT-MS method which accounts for all products derived from analyte VOCs observed in the head space during olive oil analysis. This method was subsequently used in the following two chapters.

Chapters five and six describe investigations into unfavourable and favourable sensory attributes. The most common unfavourable sensory attribute is rancidity. This attribute is accordingly analysed in more detail than the others. As sensory attributes are difficult to define in terms of chemical composition, multivariate techniques are used in these two chapters to allow comparisons between the sensory and instrumental data. The multivariate techniques used are explained in appendix A.

Chapters seven to nine outline methods commonly used to assess the antioxidant properties of olive oil. The SIFT-MS Total Oxyradical Scavenging Capacity (TOSC) assay is introduced here. This assay provides a measure of a sample's ability to scavenge radical species which are produced in the assay mixture.

Chapter eight describes the creation of a routine SIFT-MS-TOSC assay method which allows a relatively simple assay preparation and increases the sample turnover.

Chapter nine contains a comparison between the SIFT-MS-TOSC assay and some widespread methods for the analysis of common olive oil

antioxidants. An attempt to further understand the reactions occurring in the SIFT-MS-TOSC assay is also described.

**References**

(1) Visioli, F.; Galli, C. Olive Oil Phenols and Their Potential Effects on Human Health. *J. Agric. Food Chem.* **1998**, *46*, 4292-4296.

# Chapter 1

## Olive Oil Introduction

The olive tree has been known in the Mediterranean region for thousands of years, having been cultivated for its fruit a large proportion of that time. Both the olive fruit and the oil extracted from it constitute an integral part of the diet for those living in the Mediterranean, especially in Greece, Italy and Spain, the three largest producers of olives worldwide. A great deal of research has been conducted on the components of the classic 'Mediterranean diet', as it has been shown that people who eat according to this diet suffer significantly lower rates of heart disease and cancer, among other diseases(*1;2*). These people also tend to live longer and healthier lives than do those following different diets. Olive oil is the major source of fat in the 'Mediterranean diet', and there is an increasing body of research which suggests that consuming olive oil in place of other fats and oils can have a beneficial effect on human health(*3*). For this reason, there has been a popular 'swing' away from animal-based and indeed all lipids with a high saturated fat content towards olive oil and similar oils which contain more monounsaturated and polyunsaturated fats. Due in part to the discoveries made and in part to the hype which surrounds them, the consumption of olive oil is increasing in New Zealand and around the world.

To better understand olive oil, knowledge of both its production and chemical composition are necessary.

## 1.1 Olive Oil Production

The production of olive oil is a time-consuming, repetitive, yet fragile procedure. To obtain optimum oil quality and quantity, experience and patience is required. There are several steps involved, and different methods exist for performing each step. The fruit must be harvested, stored, crushed, malaxed and extracted to obtain olive oil. Each of these processes will now be described, including the various methods used at present.

### 1.1.1 Harvesting and Storage

Olives should be harvested as gently as possible, as damage to the fruit can easily lead to defects in the oil quality. If left to become overripe, they will fall off the tree and may be attacked by mould with serious consequences for the oil. In most Mediterranean countries olives are harvested when they are almost fully black(*4*). In New Zealand, especially in the South Island, the threat of frost and storms leads to earlier harvests, and a large proportion of the olives are green(*5*).

Olives are rarely able to be processed directly after harvesting, and hence must be stored. The best method of storage is in shallow plastic tubs which allow air flow over the olives and do not allow any olives to be squashed. The olives should be left in a dark, cool place such as a shed for no more than a few days before being processed. Leaving olives in sacks in the sun outside a mill for several days is sure to produce defective oil. The characteristic defect of this type of storage is called ‘fustiness’(*6*). Fustiness is described further in chapter 4.

### 1.1.2 Crushing

There are several different implements available for olive crushing, all of which perform one simple operation: they break open the olive cells and release the oil inside(*4*). Whole olives are introduced to the crusher – the stones are crushed along with the flesh of the olives(*7*). Hammer crushers are most commonly used for industrial scale olive oil production because they allow a continuous processing system. Traditionally large stone crushers were used, the largest of which weigh several tons. These are still in common use. Stone wheels are situated in a large circular bowl, and are rotated around to crush the olives. Hammer crushers generally produce oil that is richer in chlorophylls and is more bitter and spicy(*4*). Hammer crushing is commonly considered more hygienic than traditional stone crushing due to the reduced time that the oil spends in contact with the water also found in the olive fruit (oddly enough called 'vegetable water'). There is the possibility of oil fermentation by bacteria present in the vegetable water if the two phases are left in contact for an extended period of time.

In whole fruit, much of the oil is situated in vacuoles within the cells. These are broken open relatively easily to release the oil. The rest of the oil is contained in the cytoplasm of the cells(*4*). Not all oil can be removed, as some cells are left uncrushed, some oil is unable to be extracted from the pulp, and some is emulsified in the vegetable water.

Enzymes are present in olive fruit, where they decompose the oil to form VOCs. The enzymes increase in activity as ripening proceeds. Crushing permits more contact between enzymes and oil, drastically elevating the rate of VOC

production(*8*). Crushing also forms an oil-in-water emulsion, presenting a problem which is solved in the following step.

### 1.1.3 Malaxation

Once the olive cells have been crushed sufficiently to release the oil and vegetable water, the resultant paste is mixed in a process called malaxation. This step is very important because it allows the emulsified oil droplets in the paste to combine into larger drops, causing precipitation from the emulsion and increasing the effectiveness of extraction. Malaxation can improve the proportion of oil extracted from the olives by up to 10 %, typically allowing 80-85 % of the oil to be extracted from the olives(*4*). It also provides time for olive enzymes to perform the reactions necessary to produce the desired level and proportions of VOCs. The products formed at this time are largely responsible for the characteristic taste and smell of the oil. There is no set optimum duration for malaxation; it depends on the cultivar, degree of ripeness of the olives, and several other factors(*9*). There is, however, a visible change in the paste when sufficiently malaxed. The duration of malaxation is therefore controlled according to the appearance of the paste.

### 1.1.4 Extraction methods

If an oil is to be classed as ‘virgin’, it may be extracted from the olives by only mechanical means(*10*). There are several different methods of achieving this, each having its own characteristic effects on the chemistry of the oil. Other grades (e.g. ‘olive pomace oil’) may be extracted with solvents such as hexane, although they must be refined before consumption. Consequently, they are not as valuable as the virgin oils. Olive oil production equipment used in New

Zealand generally follows foreign trends, as large companies such as Pieralisi are easily able to transport equipment to New Zealand.

***Pressure***

The olive paste is spread onto mats (normally woven from a polymer such as nylon) which contain holes large enough to allow the flow of oil, yet small enough to immobilise the pulp. Several of these mats are stacked on top of each other and pressed together by machine, for approximately an hour. After this time a large proportion of the liquid has been squeezed out of the paste and collected, while a small amount remains in the pulp(*4*). The oil and vegetable water are collected together in this method – the oil is recovered in a second step using a separator. The waste from this method is dry pomace and vegetable water.

The pressure system is the most traditional form of oil extraction, although it has become more automated to take advantage of new technology. It is a batch method, meaning that once a load of pulp has been processed it must be emptied out of the machine and another load inserted. This makes the pressure method more labour intensive than centrifugation, which is described later in this chapter.

***Percolation***

Percolation takes advantage of the different surface tensions of oil and water to extract the oil from olive paste. Stainless steel blades (several thousand at a time) are dipped into the paste where they become coated with oil in preference to water. The blades are removed from the paste, the oil is drained off, and they are re-inserted into the paste. This method is not in widespread use, mostly due to the low oil yields obtained. Whenever it is used,

centrifugation is generally employed afterwards to extract the remaining oil from the pulp. The quality and yield of oil is slightly higher than that from pure centrifugation, however it is more time-consuming and requires more expensive equipment(*4*). Rather than use two methods to separate the oil from the paste, most processing plants simply use a single method and leave out percolation altogether.

***Centrifugation***

Horizontal centrifuges are the most popular method used to separate the oil from the vegetable water and pulp. The pulp is spun at high speed in a cylinder, the oil being forced to the inside due to the higher density of the water-logged pulp. Holes located at the end of the cylinder are adjusted to the correct position to capture the oil, with separate holes used to capture the vegetable water. Due to this action, they are often called 'decanters'. If the holes are in the appropriate positions, the oil is expelled from one pipe, while the waste is expelled from another(*11*).

Centrifugation is the most commonly used extraction method at present due to its speed, efficiency and the fact that it can be performed in a semi-continuous fashion, allowing it to be incorporated into a production line for olive oil extraction. Also, as the oil is in contact with the vegetable water for a shorter period of time than in other extraction methods, there is a lower probability of fermentation by micro-organisms in the water and the pulp(*12*).

## 1.2 Olive Oil Components

### 1.2.1 Triacylglycerols

95-99% of olive oil is made up of triacylglycerols (TAGs): fatty (long n-chain carboxylic) acids, normally between 12 and 22 carbon units in length,

bound to a glycerol backbone by ester bonds at their carboxyl ends. The most common triacylglycerol is that which contains three oleic acid units. The fatty acids comprising olive oil exist in well-defined relative amounts which are unique to olive oil(*13*). Olive oil also has its own unique set of TAGs, as biosynthesis in the olive fruit precludes the formation of certain TAG compositions(*13*). The fatty acid and TAG compositions of olive oil may be used to detect adulteration or alteration of the oil.

Due to the high monounsaturated fatty acid content (mostly oleic and palmitoleic), and relatively low saturated (mostly stearic and palmitic) and polyunsaturated (mostly linoleic and linolenic) fatty acid contents, olive oil is considered to be one of the best oils for human consumption(*3*). The TAGs do not, however, greatly influence the taste or smell. These sensory properties arise chiefly from the vastly less concentrated VOCs and phenolic compounds(*14*).

### 1.2.2 Volatile Organic Compounds

Over 100 different VOCs have been identified in olive oil. The majority are alcohols, aldehydes, ketones, esters or carboxylic acids, although hydrocarbons, furan derivatives, ethers, phenols, thiols, and thiophene derivatives have also been detected(*13*). The VOCs present in fresh oil are produced by the action of enzymes. One such enzyme reaction scheme is the 'lipoxygenase pathway'(*15-17*). It is suggested that the enzyme lipoxygenase oxidises the double bonds of linoleic and linolenic acids, giving rise to peroxides which are converted to $C_6$ aldehydes by several other enzymes (such as hydroperoxide lyase(*18*)). Further enzyme-catalysed reactions convert the aldehydes into the corresponding alcohols and esters. These compounds are said to be responsible for the 'green' sensory perceptions of olive oil. $C_5$ VOCs are

also formed by enzymatic action along pathways which involve hydroperoxides as intermediates(*19*). They generally exhibit much the same sensory characteristics as their $C_6$ counterparts.

It is accepted that the VOCs produced through olive biochemical channels (i.e. by enzyme catalysis) contribute to the pleasant characteristics of olive oil, and those produced by non-biological chemical means (i.e. by autoxidation, see section 3.3) contribute to the off-flavours of the oil(*12;20*). Certain VOCs (such as hexanal) are products of both processes, therefore are not reliable indicators of oil quality(*21*). VOCs (such as acetic acid) may be formed by micro-organisms growing on or in the olives. Products of processes such as this are consistently undesirable, so any metabolites from micro-organisms at levels above the sensory detection limit can have a strong effect on the oil flavour.

### 1.2.3 Phenolic Compounds

Phenolic compounds are some of the most polar compounds in olive oil(*13*), and as such their concentrations are diminished due to competitive dissolution in water during processing(*4*). The phenolic fraction is a varied mixture of many different compounds, only similar in the fact that most possess aromatic hydroxyl groups and all are sufficiently polar so as to be extracted from whole oil with a mixture of methanol and water. There are analogous methods used to isolate other classes of compounds, hence the term ‘fraction’ is often used to describe the contents of an extract. According to Boskou(*13*), phenolics are generally found in concentrations from 50 to 200 ppm in olive oil.

The phenolics in olive oil are considered to account for the bulk of the oil’s total antioxidant capacity. This is due to their radical scavenging abilities,

for the most part bestowed by the aromatic hydroxyl groups which the vast majority of these compounds contain. Salvador et. al.(*22*) calculated phenolics as having a correlation coefficient of 0.701 in relation to oxidative stability (oxidative stability is correlated with antioxidant capacity as measured by the Oxyradical Absorbance Capacity Assay(*23*)), this correlation climbing to 0.737 when considering both phenolic and α-tocopherol content(*22*). Mateos et. al.(*24*) were able to predict the oxidative stabilities of commercial olive oils by considering only the TAG composition, ortho-diphenol and α-tocopherol concentrations.

Any compounds displaying antioxidant capacity are currently and increasingly valued as food components. Olive oil has received much attention due to this trend, as has red wine, dark chocolate, blueberries and tea. However, there is currently no standard method for the measurement of antioxidant effectiveness in foods(*25;26*), and little (yet increasing(*27*)) evidence that the increased consumption of antioxidant species provides any beneficial health effects. In any case, olive oil manufacturers are understandably keen to produce olive oil with the highest possible content of any compounds considered to provide health benefits, therefore are very interested in the phenolic content of the oil they produce.

### 1.2.4 Tocopherols

Four different tocopherols have been found in olive oil and given the unimaginative names α-, β-, γ- and δ-tocopherol. Total tocopherol content in olive oil varies between 5 and 300 ppm, α-tocopherol being by far the most common, followed by β- and γ-tocopherols(*13*). δ-Tocopherol is present only in trace amounts. These compounds are known to contribute to the antioxidant

capacity of olive oil(*28*), however the nature of this contribution is not yet fully understood. Some researchers have demonstrated a synergistic relationship between the antioxidant actions of some phenolics and tocopherols(*29*).

### 1.2.5 Pigments

The green colour of unripe olives and the oil derived from them is due to chlorophyll, whereas the golden colour of more ripe oils (produced from batches containing predominantly brown or black olives) is from carotenoids(*13*).

Chlorophyll is a sensitiser which initiates oil oxidation when exposed to light by converting ground state triplet oxygen to excited state singlet oxygen(*30*). β-carotene quenches singlet oxygen(*31*), so enhances oxidative stability against light-induced (photo-) oxidation. It has also been suggested that the pigment absorbs light which would otherwise excite sensitisers, therefore reducing initiation reactions(*32*). The effect of β-carotene during oxidation in the dark – where reactions are not initiated by pigment sensitisation – depends on the conditions under which the reactions occur. Whether β-carotene is an antioxidant or a prooxidant and how effective it is, depends on the concentration of both itself and oxygen(*33*), as well as the chemical environment(*34*).

### 1.2.6 Peroxides

The peroxides (termed ‘primary oxidation products’) found in olive oil are produced via oxidation of fatty acids(*35*). Peroxides are formed at low concentrations under controlled conditions by the lipoxygenase enzyme. Further enzymatic reactions are performed, eventually producing VOCs as mentioned previously(*17*). However, oxidation of triacylglycerols is not the sole reserve of enzymes – less specific peroxide formation by excited singlet oxygen may also occur. When these peroxides break down they also form VOCs (‘secondary

oxidation products'), however these VOCs are often unpleasant and contribute to the sensory defect known as rancidity(*12*).

### 1.2.7 Squalene

Squalene is one of only two hydrocarbons present in appreciable amounts in olive oil, the other being β-carotene. It is a major component (~40%) of the unsaponifiable fraction, the remaining material after addition of an alkaline hydroxide, extraction with a solvent (such as diethyl ether), and expulsion of volatiles at 103 °C. Saponification essentially removes all fatty acids, glycerol-bound or otherwise, along with all VOCs(*13*). Levels of squalene (a sterol precursor) in the body achieved by including olive oil in the diet (around 40 g per day, a common value for people in Mediterranean countries) may have an inhibitory effect on cancer development(*36*), although current evidence is not overwhelming and this may be solely due to other olive oil constituents, such as the abundant phenolic antioxidants.

## 1.3 Olive Oil Testing

Not all olive oil is fit for human consumption directly after processing due to contamination or spoilage. One of several factors may be responsible. Whatever the reason, this oil needs to be detected and diverted from sale. Three standard tests must be passed by all olive oil before it may be offered for sale in International Olive Oil Council (IOOC) member countries. These tests are outlined below.

### 1.3.1 Free Acidity

If olive oil is left in contact with water, a significant degree of acid-catalysed hydrolysis(*37*) may occur. In this event, di- and mono-acylglycerols and free fatty acids (FFAs) are formed(*13*). These are all considered undesirable

components because they result from oil degradation and in the case of FFAs, alter the pH - seriously affecting the quality of the oil. As FFAs alter the pH of the oil, a titration may be used to determine their concentrations, and has been adopted as an official test by both the IOOC and the International Union of Pure and Applied Chemistry(*38*). The FFA content is used as one of the distinctions between olive oil grades ('extra virgin' must have less than 0.8% FFA(*10*), expressed as 'free oleic acid', sometimes confusingly referred to as simply 'oleic acid'), so is a very important and closely monitored quantity.

### 1.3.2 Peroxide Value

The peroxide value (PV) test measures primary oxidation products, which are peroxides as described above. Peroxide concentration in olive oil is commonly measured in milliequivalents of active oxygen per kilogram ($meqO_2\ kg^{-1}$)(*13*). This is one of the main factors in determining the class of an olive oil, as it gives an indication of the extent of oil oxidation. For an oil to be declared 'extra virgin', the IOOC states that it must contain no more than 20 $meqO_2\ kg^{-1}$(*10*), usually determined by titration with sodium thiosulfate solution after treatment with potassium iodide(*38*). Fresh olive oil should have a PV of less than 4 $meqO_2\ kg^{-1}$ (personal communication, Margaret Edwards, Olives New Zealand sensory panel head), and Olives New Zealand (ONZ, the government-recognised group of olive growers and olive oil producers in New Zealand) will not accredit as 'extra virgin' any oil with a PV of more than 15 $meqO_2\ kg^{-1}$(*39*).

### 1.3.3 The Sensory Panel Test

Many of the VOCs identified in olive oil exhibit their own characteristic odour or taste when smelled or tasted, respectively. The sensations induced by

these compounds are assigned various descriptors comparing them with commonly encountered tastes or smells for the purpose of classifying olive oils by quality. Descriptors may be either desirable (such as fruity, bitter and pungent) or undesirable (such as winey, musty and metallic)(*12*).

The analysis of olive oil by sensory panels is closely monitored by the International Olive Oil Council (IOOC)(*40*). The IOOC accredits panels of judges it deems worthy to award oils with the titles of 'extra virgin' or 'virgin'. These judges must undergo rigorous training and testing, showing discrimination and repeatability in the perception of tastes and odours. After receiving accreditation, a panel is continually tested to ensure it remains at the level of performance required by the IOOC.

Each class of oil (e.g. extra virgin, virgin, etc.) has assigned minimum scores for positive attributes and maximum scores for negative attributes as detected by the sensory panel. For example, if an oil is to be classed as 'extra virgin', it must have no detectable defects (median zero), and a noticeable 'fruity' attribute (median greater than or equal to 2.5 on a scale of ten)(*41*). For an olive oil to be recognised as extra virgin in IOOC member countries, it must have passed the appropriate tests, which include the sensory test carried out by an accredited panel. This testing method may seem very open to interpretation, however achieving IOOC accreditation is a lengthy and rigorous process (personal communication, Margaret Edwards, Olives New Zealand sensory panel head). Therefore once a panel has been operating for a time sufficient to be accredited by the IOOC, each panel member knows which sensations to look for when tasting and smelling olive oil, and is able to quantify each consistently. A minimum amount of agreement is necessary between members of the panel

for the judgement to be accepted (no more than 20% robust standard deviation is accepted (*41*)), allowing inconsistent measurements to be identified and repeated to gain a more precise assessment. 'Robust' standard deviation is used to describe the variation in tasting data as it takes into account the possibility of outliers and excludes extreme data from the calculation.

The subjectivity of sensory testing is most pronounced when grading oils for competition. Competition judging is another facet of olive oil sensory analysis. It is not regulated as is the official sensory test, and the results do not decide whether an oil is 'lampante' or 'extra virgin'. All oils considered for competition judging must already be certified 'extra virgin' class – competition judging awards gold, silver or bronze medals from whichever organisation which runs the awards. Different organisations may hold their own awards. For example, the IOOC each year holds the 'Mario Solinas Quality Award'(*42*). Competition judging is very much based on the taster's opinion of the whole oil, the overall impression outweighing the contribution from individual flavour characteristics. No information about the producers, processors, specific geographic origin or even the colour of the oil is made known to the tasters, so the only opinion possible is based on the perceived odour and taste. As the entire process is in place to assess the potential for enjoyment of the oil, a judging technique which is biased by enjoyment is perfectly legitimate, even desirable. The ultimate goal of instrumental analysis is to replicate the results of competition judging, as many subtleties are factored into this which are not included in the official sensory test. However, as competition judging is not as firmly regulated as the official sensory test, there is variation between panels

such that forming a consensus of the results of different panels may be just as difficult as explaining the results in terms of the chemical species responsible.

## 1.4 References


(1) Keys, A. Mediterranean Diet and Public Health: Personal Reflections. *Am. J. Clinical Nutr.* **1995**, *61*, 1321S-1323S.

(2) Wahrburg, U.; Kratz, M.; Cullen, P. Mediterranean Diet, Olive Oil and Health. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 698-705.

(3) Beardsell, D.; Francis, J.; Ridley, D.; Robards, K. Health Promoting Constituents in Plant Derived Edible Oils. *J. Food Lipids* **2002**, *9*, 1-34.

(4) Di Giovacchino, L. Olive Harvesting and Olive Oil Extraction. In *Olive Oil: Chemistry and Technology*; D. Boskou, Ed.; AOCS Press: Champaign, IL, USA, **1996**.

(5) Dahl, P., Personal Communication from Local Olive Oil Producer, 2005.

(6) Angerosa, F.; Di Giacinto, L.; Solinas, M. Influenza dello Stoccaggio in Massa delle Olive sull'Aroma degli Oli di Risulta: Valutazione del Difetto di "Riscaldo" Mediante Analisi HPLC e GLC dei Componenti Volatili. *Riv. Merceol.* **1990**, *29*, 275-294.

(7) Mulinacci, N.; Giaccherini, C.; Innocenti, M.; Romani, A.; Vincieri, F. F.; Marotta, F.; Mattei, A. Analysis of extra virgin olive oils from stoned olives. *J. Sci. Food Agric.* **2005**, *85*, 662-670.

(8) Kalua, C.; Allen, M.; Bedgood, D. J.; Bishop, A.; Prenzler, P.; Robards, K. Olive Oil Volatile Compounds, Flavour Development and Quality: A Critical Review. *Food Chem.* **2007**, *100*, 273-286.

(9) Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds and Phenolic Composition of Virgin Olive Oil: Optimization of Temperature and Time of Exposure of Olive Pastes to Air Contact during the Mechanical Extraction Process. *J. Agric. Food Chem.* **2003**, *51*, 7980-7988.

(10) International Olive Oil Council COI/T.15/NC no.3 Trade standard applying to olive oil and olive-pomace oil 2003. www.internationaloliveoil.org/chemistry2.asp (accessed Nov 8, 2006)

(11) The Olive Oil Source website. Olive Presses and Mills Explanation. http://www.oliveoilsource.com/mill_and_press_facts3.htm (accessed 12 March, 2007)

(12) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 639-660.

(13) Boskou, D. *Olive Oil: Chemistry and Technology*; AOCS Press: Champaign, IL, USA, 1996.

(14) Aparicio, R.; Morales, M. T.; Luna, G.; Aparicio-Ruiz, R. Biochemistry and Chemistry of Volatile Compounds Affecting Consumers' Attitudes towards Virgin Olive Oil. In *Proc. Phytochem. Soc. Eur.*, **2000**; pp 3-14.

(15) Angerosa, F.; d'Alessandro, N.; Basti, C.; Vito, R. Biogeneration of Volatile Compounds in Virgin Olive Oil: Their Evolution in Relation to Malaxation Time. *J. Agric. Food Chem.* **1998**, *46*, 2940-2944.

(16) Feussner, I.; Wasternack, C. The Lipoxygenase Pathway. *Annu. Rev. Plant Biol.* **2002**, *53*, 275-297.

(17) Sánchez, J.; Harwood, J. L. Biosynthesis of Triacylglycerols and Volatiles in Olives. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 564-573.

(18) Vick, B. A.; Zimmerman, D. C. Lipoxygenase and Hydroperoxide Lyase in Germinating Watermelon Seedlings. *Plant Physiol.* **1976**, *57*, 780-788.

(19) Vichi, S.; Pizzale, L.; Conte, L. S.; Buxaderas, S.; López-Tamames, E. Solid-Phase Microextraction in the Analysis of Virgin Olive Oil Volatile Fraction: Characterization of Virgin Olive Oils from Two Distinct Geographical Areas of Northern Italy. *J. Agric. Food Chem.* **2003**, *51*, 6572-6577.

(20) Angerosa, F.; Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds in Virgin Olive Oil: Occurrence and Their Relationship With the Quality. *J. Chrom. A* **2004**, *1054*, 17-31.

(21) Morales, M. T.; Luna, G.; Aparicio, R. Sensory and Chemical Evaluation of Winey-Vinegary Defect in Virgin Olive Oils. *Eur. Food Res. Technol.* **2000**, *211*, 222-228.

(22) Salvador, M.; Aranda, F.; Fregapane, G. Contribution of Chemical Components of Cornicabra Virgin Olive Oils to Oxidative Stability. A Study of Three Successive Crop Seasons. *J. Amer. Oil Chem. Soc.* **1999**, *76*, 427-432.

(23) Ninfali, P.; Bacchiocca, M.; Biagiotti, E.; Servili, M.; Montedoro, G. Validation of the Oxygen Radical Absorbance Capacity (ORAC) Parameter as a New Index of Quality and Stability of Virgin Olive Oil. *J. Amer. Oil Chem. Soc.* **2002**, *79*, 977-982.

(24) Mateos, R.; Trujillo, M.; Pérez-Camino, C.; Moreda, W.; Cert, A. Relationships between Oxidative Stability, Triacylglycerol Composition, and Antioxidant Content in Olive Oil Matrices. *J. Agric. Food Chem.* **2005**, *53*, 5766-5771.

(25) Frankel, E. N.; Meyer, A. S. The Problems of Using One-Dimensional Methods to Evaluate Multifunctional Food and Biological Antioxidants. *J. Sci. Food Agric.* **2000**, *80*, 1925-1941.

(26) Prior, R.; Wu, X.; Schaich, K. Standardized Methods for the Determination of Antioxidant Capacity and Phenolics in Foods and Dietary Supplements. *J. Agric. Food Chem.* **2005**, *53*, 4290-4302.

(27) Vissers, M. N.; Zock, P. L.; Roodenburg, A. J.; Leenen, R.; Katan, M. B. Olive Oil Phenols Are Absorbed in Humans. *J. Nutr.* **2002**.

(28) Deiana, M.; Rosa, A.; Cao, C. F.; Pirisi, F. M.; Bandino, G.; Dessì, A. Novel Approach to Study Oxidative Stability of Extra Virgin Olive Oils: Importance of α-Tocopherol Concentration. *J. Agric. Food Chem.* **2002**, *50*, 4342-4346.

(29) Hudson, B.; Lewis, J. Polyhydroxy Flavonoid Antioxidants for Edible Oils. Structural Criteria for Activity. *Food Chem.* **1983**, *10*, 47-55.

(30) Küpper, H.; Dedic, R.; Svoboda, A.; Hála, J.; Kroneck, P. M. Kinetics and Efficency of Excitation Energy Transfer From Chlorophylls, Their Heavy Metal-Substituted Derivatives, and Pheophytins to Singlet Oxygen. *Biochim. Biophys. Acta* **2002**, *1572*, 107-113.

(31) Foote, C. S.; Denny, R. W. Chemistry of Singlet Oxygen VII. Quenching by β-Carotene. *J. Amer. Chem. Soc.* **1968**, *90*, 6233-6234.

(32) Hansen, E.; Skibsted, L. H. Light-Induced Oxidative Changes in a Model Dairy Spread. Wavelength Dependence of Quantum Yields. *J. Agric. Food Chem.* **2000**, *48*, 3090-3094.

(33) Burton, G.; Ingold, K. β-Carotene: An Unusual Type of Lipid Antioxidant. *Science* **1984**, *224*, 569(565).

(34) Yanishlieva, N. V.; Aitzetmüller, K.; Raneva, V. G. β-Carotene and Lipid Oxidation. *Fett/Lipid* **1998**, *100*, 444-462.

(35) Belitz, H.-D.; Grosch, W. *Food Chemistry*, 2nd English ed.; Springer-Verlag: Berlin, Germany, 1999.

(36) Newmark, H. L. Squalene, Olive Oil, and Cancer Risk: Review and Hypothesis. *Ann. N. Y. Acad. Sci.* **1999**, *889*, 193-203.
(37) Loudon, G. M. *Organic Chemistry*, 3rd ed.; The Benjamin/Cummings Publishing Company, Inc.: Redwood City, CA, USA, 1995.
(38) Tsimidou, M.; Boskou, D. Olive Oil Analysis. In *Olive Oil: Chemistry and Technology*; D. Boskou, Ed.; AOCS Press: Champaign, IL, USA, **1996**.
(39) Olives New Zealand website. Extra Virgin Olive Oil 2006 Certification Programme. www.olivesnz.org.nz/certification.cfm (accessed Nov 8, 2006)
(40) International Olive Oil Council COI/T.20/Doc. no. 13/Rev. 1 General Methodology for the Organoleptic Assessment of Virgin Olive Oil. http://www.internationaloliveoil.org/downloads/orga4.pdf (accessed Mar 22, 2007)
(41) International Olive Oil Council COI/T.20/Doc. no. 15/Rev. 1 Sensory Analysis of Olive Oil Method: Organoleptic Assessment of Virgin Olive Oil. http://www.internationaloliveoil.org/downloads/orga6.pdf (accessed Feb 13, 2007)
(42) International Olive Oil Council Mario Solinas Quality Award. http://www.internationaloliveoil.org/chemistry5.asp (accessed 22 Mar, 2007)

# Chapter 2

## The Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS) Technique

### 2.1 Instrumentation

#### 2.1.1 Introduction

Selected Ion Flow Tube (SIFT) instruments were first constructed in the late 1970s(*1*) as an extension of the flowing afterglow (FA) technique that was used to study the kinetics of reactions of neutral molecules in the 1960s. The FA technique was first applied to the investigation of ion-molecule reactions by Ferguson et. al. in the late 1960s(*2*). The initial use of FA and SIFT technology was in modelling reactions between gas phase ions and neutral species under outer atmospheric and interstellar conditions. During the 1990s the SIFT instrument's extreme sensitivity facilitated its extension to the monitoring of analyte concentrations in air(*3*). This variation of SIFT is known as Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS). SIFT-MS originated from the observation that several easily formed ionic species generated from air were unreactive with the major constituents of ambient air, yet reacted with many minor components to form characteristic products. Many reactions proceeded via only one or two pathways, with predictable reaction rates and branching ratios. Thus, the SIFT-MS technique was born. SIFT-MS is ideal for analysis and monitoring of gas samples (including whole-air) which contain analyte compounds of interest down to several parts per billion by mole (ppb/mol). By

altering the flow rate of sample gas, there is no upper limit of concentration for the quantitation of sample components(*3*).

Each SIFT-MS instrument can be divided into four operating regions: the ion creation region, the ion selection region, the reaction region and the ion detection region. They are arranged as shown in figure 2.1. The function of each region is described in sections 2.1.2 to 2.1.5.

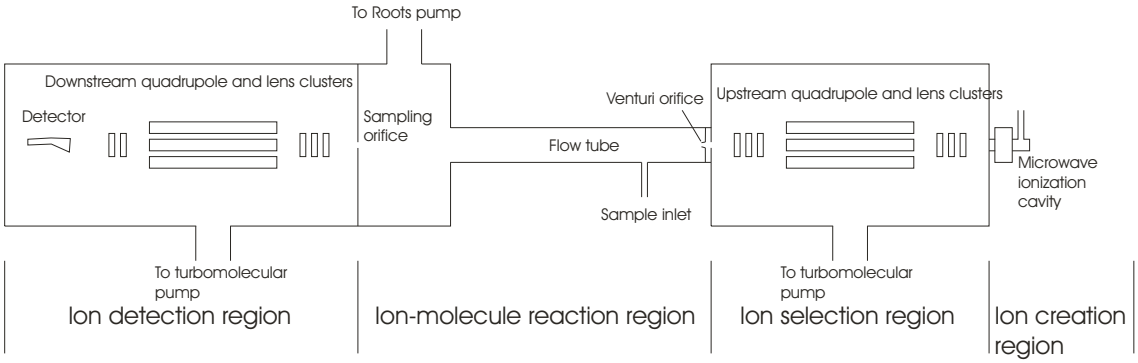



**Figure 2.1.**, a representation of the SIFT-MS instrument used in the present research. The direction of gas and ion flow is from right to left.

### 2.1.2 The Ion Creation Region

There are three reagent ions commonly used in SIFT-MS analysis: $H_3O^+$, $NO^+$ and $O_2^+$(*4*). These are the three terminal positive ions produced from the break down of moist air when exposed to sufficiently high intensity microwave energy. Atmospheric air is drawn through a sample of distilled water and into a discharge chamber, where it is exposed to microwave energy and becomes a plasma. Heating generates ions and these are accompanied by electronic transitions in the visible

region of the spectrum. An electrostatic lens extracts positive ions from the plasma and transfers them to the ion selection region.

### 2.1.3 The Ion Selection Region

The positive ions extracted from the ion creation region are focussed by an Einzel lens arrangement into a quadrupole mass filter. The quadrupole mass filter selects the desired reagent ion, and another lens arrangement focuses the ions into the flow tube. The quadrupole mass filter is contained in its own chamber which is maintained at medium vacuum with a pressure of approximately $4.7 \times 10^{-3}$ Pa ($3.5 \times 10^{-5}$ Torr) by a turbomolecular pump backed by a rotary pump.

### 2.1.4 The Reaction Region

The mass-selected reagent ions exit the selection region through a small orifice and enter the flow tube. The ions are carried along the flow tube reactor in a stream of helium and argon gases which are injected through a Venturi inlet in a circular pattern through two annuli. Helium is introduced through the inner annulus and this annulus surrounds the ion injection orifice. The argon gas is introduced through an outer annulus. The carrier gases slow radial diffusion of the ions (see 'Mass Discrimination' below). The flow directly after the Venturi inlet is turbulent, but assumes laminar flow after a short distance. The analyte (often as the head space above a liquid) is introduced into the flow tube in the region of laminar flow. As soon as the analyte is introduced into the flow tube, it begins reacting with the reagent ions. Typical ion/molecule reactions proceed at or near the collision rate(*4*) and have negligible activation energies.

The flow tube is under ‘steady-state’ conditions before the addition of a sample, with a constant population of reagent ions ready to react when a sample is introduced downstream from the ion entrance orifice. The most convenient samples for SIFT-MS analysis are those which also allow approximate ‘steady-state’ conditions while sampling. A stream of sample maintained at a constant flow rate permits many sampling points to be taken, leading to very good measurement precision. A sample in which the concentrations of analytes are changing may also produce excellent results, although the sample should still ideally be in the form of a stream with a constant flow of gas. Difficulties may arise when the sample is not able to be introduced in this way, leading to compromises as sometimes encountered in the analysis of breath(*3*) or the head space of small bottles(*5*). Sampling time is limited for these types of analyses, making investigation of the analytes more difficult than for ideal sample types.

At the end of the reaction region is another small hole, called the ‘downstream sampling orifice’. This is on the axis of the flow tube reactor. The downstream sampling orifice permits the ions entry into the detection region. The more ions that enter through the orifice, the more are available for detection and the more sensitive the subsequent measurement. This is why it is so important to reduce radial ion diffusion.

A Roots pump backed by a rotary pump is used to evacuate the reaction region. A Roots pump is needed here, as the reaction region has the highest flow of gas of any region in the instrument (approximately 70 Torr L $s^{-1}$). A typical flow tube pressure is 0.80 Torr.

The time taken for the analyte molecules to travel from the point at which they enter the flow tube to the downstream sampling orifice is calculated for each sampling point from known and measured instrument parameters, and is normally approximately 3.4 ms. This gives an indication of the potential response time of a SIFT-MS instrument. During this time, called the 'reaction time', the analyte molecules are free to react with the reagent ions(*6*). Once they enter through the downstream sampling orifice, however, the pressure is much lower (1 x $10^{-5}$ Torr), collisions are correspondingly fewer and the reaction is halted.

### 2.1.5 The Ion Detection Region

Immediately past the downstream sampling orifice is another set of three lenses forming an Einzel lens which focuses ions into the downstream quadrupole mass filter. This quadrupole mass filter provides the only separation experienced by the product ions in the SIFT-MS technique. The pressure in the detection region is approximately five orders of magnitude lower than that in the reaction region, so the number of collisions and therefore the reaction rate is greatly reduced. For this reason, the ion/molecule reactions can be assumed to cease once the ions pass the downstream sampling orifice.

Once through the quadrupole mass filter, the mass-selected ions are detected by a particle multiplier, and the current measured in the chosen time interval is converted into counts per second. The variable time interval over which the ions are counted lends a great deal of flexibility to the SIFT-MS technique. This interval controls the balance between the response time of the instrument to changes in analyte concentration and the precision of the ion counting. A smaller counting

interval means the number of ions counted is reported more quickly and more often, while a larger interval means that a larger sample of ions is taken, giving higher precision for that interval. Counting fewer ions per interval yet having more repetitions means the same number of ions is counted overall, so both scenarios will give equal precision.

The counting of each ion is a discrete event and the count rate is approximated by the Poisson distribution(*7*). The expected standard deviation of a variable which follows the Poisson distribution is the square root of the mean number of ions counted during a given interval. The observed precision of SIFT-MS measurements are often very close to this value. It is necessary to take the measurement precision into account when defining counting intervals in the development of SIFT-MS methods to ensure that acceptable precision is achieved.

There are two different methods of data acquisition possible using a SIFT-MS instrument. They are named mass scans and Selected Ion Monitoring (SIM) scans. A mass scan may be thought of as a ‘snapshot’ of the sample, where all ions formed between a specified m/z (mass to charge ratio: ‘m’ is mass, ‘z’ is charge) range are counted in turn. As there are often at least 150 different m/z values to be analysed, only a small amount of time may be spent counting each one (typically 10-20 ms), and the results obtained are not suitable for quantitative purposes. Mass scans are excellent for primary investigation of unknown samples, as all products formed are detected. When mass scans with all three precursor ions are performed on a sample, a full data set is obtained which allows the identification of the VOCs present in the sample. It must, however, be assured that the sample does not change

noticeably over the course of the mass scan. The most disruptive change referred to here is caused by reduced sample flow due to reducing pressure inside a fixed-volume sampling vessel. Any fixed-volume sampling vessel must be sufficiently large so that pressure changes are negligible, as a changing sample may produce some very confusing spectra.

A SIM scan is used to provide quantitative information about the VOCs present in a sample. SIM scans are very useful for monitoring changes in a sample over time or detecting differences between samples. A selection of m/z values is scanned in turn (each for an individually specified length of time), and this is repeated for as long as required. As fewer m/z values are involved, each response may be counted for longer than in a mass scan (providing greater sensitivity and precision) and the time taken to complete each scan cycle is less. A cycle time of five seconds is not uncommon, although they may vary widely. Hence, SIM scans provide adjustable time resolution to monitor changes in VOC concentrations in a sample. The use of SIM scans is generally confined to the analysis of known samples. The identification of VOCs present in a sample prior to the use of a SIM scan is necessary due to the possibility of isobaric product ions. However, once a type of sample has been characterised, a SIM scan allows the rapid collection of vast amounts of quantitative data.

As described above, the precision of a SIFT-MS measurement depends on the number of ions counted during the relevant counting intervals. This may be affected by several different factors; the correct solution is the one that gives the most acceptable trade-off between them. The major factors to consider are the

precision and sensitivity (these two are closely related), the number of analytes being measured, the response time and the detection limit. The first three are to be maximised, while the last two should be minimised. Incorrect values for any of these factors may cause less than optimal results to be obtained from the SIM method. The aim is to choose method parameters which give optimum values for these factors which are far from these critical values. As the measurement of appropriate values for some of these factors (such as response time) are more empirical than those of others (such as precision), several different versions of a new method with different numbers of analytes, counting intervals etc. should be trialled to determine the most satisfactory trade-off position.

## 2.2 Interpreting the Results

### 2.2.1 Calculating Analyte Concentration

The concentration of an analyte in the flow tube at a give time is calculated via equation 2.1 (from Španěl and Smith (*3*)):

$$c(M) = \frac{I_p}{t_r k I_i} \qquad (2.1)$$

where c(M) is the concentration of analyte M in the flow tube in molecules $cm^{-3}$ (often referred to as the 'number density'), $I_p$ is the number of counts per second of product ion detected, $t_r$ is the time available for reaction between the reagent ions and the analyte, k is the rate constant for reaction between reagent ion and analyte in $cm^3$ $molecule^{-1}$ $s^{-1}$ and $I_i$ is the number of counts per

second of reagent ion detected. A more realistic situation, where multiple reagent ions (e.g. water clusters of the primary reagent ion) exist with different reactivities, more than one product ion is produced from the analyte and the effects of mass discrimination are factored in is represented by the approximation given by Španěl, Dryahina and Smith(*6*) (equation 2.2):

$$c(M) = \frac{\frac{I_{p1}}{D_{fp1}} + \frac{I_{p2}}{D_{fp2}} + \cdots}{t_r\left(f_{i1}I_{i1}k_1 + \frac{f_{i2}I_{i2}\frac{k_1+k_2}{2}}{D_{fi2}} + \cdots\right)} \qquad (2.2)$$

where $I_{pm}$ is the number of counts per second of product ion m detected (many compounds give multiple product ions), $D_{fpm}$ is a correction for diffusion enhancement of product ion m (see 'Mass Discrimination' below), $f_{in}$ is a correction applied to account for conversion of reagent ions into their n clustered species, $I_{in}$ is the number of counts per second of reagent ion n detected, $k_n$ is the rate constant for reaction of reagent ion n with the analyte and $D_{fin}$ is a correction for diffusion enhancement of reagent ion n. This $D_{fin}$ term arises because the diffusive loss of heavier ions is usually less than that of lower mass ions. The approximation in equation 2.2 holds when there is a sufficiently low number density of analyte so that no more than 10 % of the reagent ions are lost through reaction. The reaction time $t_r$ (in seconds) is calculated from the flow dynamics of the carrier gases according to equation 2.3:

$$t_r = \frac{P_{ft}V}{R_v(\Phi_c + \Phi_s)} \qquad (2.3)$$

where $P_{ft}$ is the pressure in the flow tube (measured in Torr), V is the internal volume of the flow tube (in L), where V = Al (A is the internal cross-section of the flow tube and l is the distance between the point of sample introduction and the downstream sampling orifice). $R_v$ is a dimensionless factor introduced to describe the velocity of the ions relative to the carrier gas, $\Phi_c$ is the total flow rate of carrier gas in the flow tube (in Torr L $s^{-1}$) and $\Phi_s$ is the flow of sample gas into the flow tube (also in Torr L $s^{-1}$).

Once the number density of analyte in the flow tube is known, the concentration of analyte in the original sample mixture may be calculated using equation 2.4(*3*):

$$\text{ppb analyte} = \frac{c(M) \times 10^3 RT(\Phi_c + \Phi_s)}{N_A P_{ft} \Phi_s} \times 10^9 \qquad (2.4)$$

where R is the gas constant (62.36 L Torr $K^{-1}$ $mol^{-1}$), T is the temperature (in K) and $N_A$ is Avogadro's number, $6.02 \times 10^{23}$. The number density c(M) is converted to molecules $L^{-1}$ by multiplying by $10^3$. The terms represented by Roman letters in equation 2.4 give the proportion of analyte molecules in the flow tube gas. The proportion is then multiplied by the dilution factor of the sample in the flow tube (the total gas flow divided by the sample gas flow) to give the proportion of

analyte which was admitted in the original sample gas. The proportion of analyte in the sample gas is then multiplied by $10^9$ to give the number of parts per billion of analyte by mole.

### 2.2.2 Mass Discrimination

An added complication presents itself in the detection of ions of different masses. Ions of lighter mass are more easily transmitted through a quadrupole mass filter and detected by an electron multiplier than those of higher mass. This occurs in all quadrupole-based mass spectrometric systems, but is not a problem for most other techniques, as quantification is either not necessary or is made empirically by direct and frequent comparison with standards of known concentration. By contrast, SIFT-MS provides online quantitation without constant calibration, therefore the discrimination must be determined for each instrument to obtain accurate results. Španěl and Smith(*6*) have described a method for determining mass discrimination ($M_r$). This involves complete conversion of reagent ions with selected analytes to form product ions of a single mass. $H_3O^+$ at m/z = 19 is defined as having a $M_r$ value of 1, and all other masses are measured relative to this. A selection of analytes which give product ions of different masses is used to enable the generation of a discrimination function which can be interpolated to account for ions of any mass.

The downstream sampling orifice is situated in a disc, the current on which is able to be measured. Any current detected on this disc is due to ions from the flow tube colliding with it. In this way, it gives a measure of the number of ions arriving at the end of the flow tube. Comparison of the current on the disc with the

counts arriving at the detector is used to determine the difference in transmission of ions of different mass through the quadrupole chamber. $H_3O^+$ ions are selected and flow down the flow tube, where the current produced by them on the downstream disc ($I_0$) is measured. Those ions which pass through the downstream sampling orifice and travel through the quadrupole mass filter are counted by the detector (giving a value $C_0$). The $H_3O^+$ ions are then converted fully to another species of different mass by the addition of a large amount of a single analyte into the flow tube. The current in this situation is also measured on the downstream disc ($I_d$), and the number of ions reaching the detector are measured ($C_d$). The current of the chosen analyte ions is divided by the current of the reagent ions arriving at the downstream disc. This is multiplied by the number of reagent ions divided by the analyte product ions reaching the detector, as in equation 2.5. If the two fractions are reciprocal, the transmission is identical for both species and there is no mass discrimination displayed through the quadrupole chamber. This produces a $M_r$ value of 1. In most cases product ions from analytes of higher mass than $H_3O^+$ will be lost more readily through the downstream quadrupole chamber, and the fraction $I_d/C_d$ will be larger than $I_0/C_0$, giving a $M_r$ value greater than 1.

$$M_r = \frac{I_d}{I_0}\frac{C_0}{C_d} \qquad (2.5)$$

In SIFT-MS however, this is not the complete story. Differential transmission of ions is not confined to the quadrupole chamber. An additional mass

discrimination effect acts in the flow tube before the ions even reach the detection region.

Ions diffuse off axis in the flow tube as mentioned previously, but this effect is not equal for all ions. Ions of higher mass do not diffuse as quickly as the lighter ions – this is why helium is used, as it slows the radial diffusion of lighter ions. Intermediate-sized ions can also diffuse noticeably off axis, so argon is also used to slow this diffusion. The amount of argon in the instrument used for the experiments described here is just over one sixth that of helium (9.3 compared with 57.6 Torr L $s^{-1}$). These carrier gases help, but they only slow the diffusion, they do not prevent it. The ion loss due to differential diffusion must also be measured to obtain accurate results.

The same strategy is used to measure diffusive discrimination as quadrupole mass filter discrimination. On admission of an air sample containing a high concentration of the chosen analyte, the analyte molecules are assumed to react instantly with all available reagent ions, producing the same number density of product ions on introduction of the analyte as there were reagent ions when the analyte was absent. This means that only the difference in diffusion of the two ionic species must be considered. According to Španěl and Smith(*8*), the diffusion enhancement ($D_e$) is described by equation 2.6:

$$D_e = \frac{e^{\left(\frac{D_p\left(H_3O^+\right) - D_p\left(MH^+\right)}{\Lambda^2} t_r\right)^{-1}}}{\frac{D_p\left(H_3O^+\right) - D_p\left(MH^+\right)}{\Lambda^2} t_r} \qquad (2.6)$$

where $D_p(X^+)$ is the free diffusion coefficient for species $X^+$, $\Lambda$ is a term which describes radial diffusion of ions, and $t_r$ is the reaction time (or the time taken for the ions to travel from the sample inlet to the downstream sampling orifice).

Conveniently, the numerator in the above equation may be rewritten, due to the following relationship (equation 2.7(*8*)):

$$\frac{I_d}{I_0} = e^{\left(\frac{D_p\left(H_3O^+\right) - D_p\left(MH^+\right)}{\Lambda^2} t_r\right)} \qquad (2.7)$$

So equation 2.6 may be rewritten as:

$$D_e = \frac{\frac{I_d}{I_0} - 1}{\ln\left(\frac{I_d}{I_0}\right)} \qquad (2.8)$$

$I_d$ and $I_0$ have already been measured to account for the quadrupole mass discrimination, so no further measurements are needed to determine $D_e$.

$D_f$ is the full SIFT-MS discrimination factor for a given mass, and is calculated by equation 2.9(*8*):

$$D_f = \frac{M_r}{D_e} \tag{2.9}$$

$D_f$ values are included in the calculation of concentrations as shown in equation 2.2.

The process described here allows the calculation of analyte concentrations from experimentally measured values.

## 2.3 References

(1) Adams, N.; Smith, D. The Selected Ion Flow Tube (SIFT): A Technique for Studying Ion-Neutral Reactions. *Int. J. Mass Spectrom. Ion Phys.* **1976**, *21*, 349-359.

(2) Ferguson, E.; Fehsenfeld, F.; Schmeltekopf, A. Flowing Afterglow Measurements of Ion-Neutral Reactions. *Adv. Atom. Mol. Phys.* **1969**, *5*, 1-56.

(3) Španel, P.; Smith, D. Selected Ion Flow Tube: A Technique for Quantitative Trace Gas Analysis of Air and Breath. *Med. & Biol. Eng. & Comput.* **1996**, *34*, 409-419.

(4) Smith, D.; Španel, P. Selected Ion Flow Tube Mass Spectrometry (SIFT-MS) for On-Line Trace Gas Analysis. *Mass Spectrom. Rev.* **2005**, *24*, 661-700.

(5) Španel, P.; Diskin, A. M.; Abbott, S. M.; Wang, T.; Smith, D. Quantification of Volatile Compounds in the Headspace of Aqueous Liquids Using Selected Ion Flow Tube Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 2148-2153.

(6) Španel, P.; Dryahina, K.; Smith, D. A General Method for the Calculation of Absolute Trace Gas Concentrations in Air and Breath from Selected Ion Flow Tube Mass Spectrometry Data. *Int. J. Mass Spectrom.* **2006**, *249-250*, 230-239.

(7) Box, G. E.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*; John Wiley & Sons: New York, NY, USA, 1978.

(8) Španel, P.; Smith, D. Quantitative Selected Ion Flow Tube Mass Spectrometry: The Influence of Ionic Diffusion and Mass Discrimination. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 863-872.

# Chapter 3

## Olive Oil Volatile Organic Compounds

### 3.1 VOC Analysis Techniques

The first reported instrumental analyses of olive oil VOCs were performed in the late 1960s. Since then several important advances have been made, particularly in instrumental sensitivity and sampling techniques. Gas chromatography was, and still is, the standard technique for analysis of VOCs, although several different techniques have been introduced recently.

#### 3.1.1 Chromatographic Analysis

Direct Gas Chromatography (GC) of olive oil head space VOCs is not normally performed. This is because the concentrations of the majority of VOCs in olive oil are too low to allow detection without an enrichment step(*1*). The first GC experiments on olive oil met with some success, however the identified compounds reported are relatively few and do not resemble closely those reported today using modern instrumentation and techniques. The first successful identification of a large number of VOCs present in olive oil was performed by Flath, Forrey and Guadagni in 1973(*2*). The full list of identified compounds numbers 77, and is still a very comprehensive list when compared with research conducted today.

Adsorption techniques are now widely used, and may be divided into two categories: Dynamic Head Space (DHS) and Static Head Space (SHS) techniques. In DHS techniques, VOCs are stripped from olive oil by a stream of nitrogen gas and deposited on a fine powder adsorbent such as activated charcoal or Tenax. The VOCs are then desorbed at elevated temperature and

measured. SHS techniques are more convenient as no pump or gas supply is necessary, however the amount of sample extracted is also less. Solid-Phase Microextraction (SPME) is the most popular SHS enrichment technique. SPME involves the use of a thin fibre which is exposed to the head space in a vial for a certain length of time, after which it is sealed in a holder. The VOCs adsorbed on the fibre are then swept onto a GC column at elevated temperature(*1*).

These new techniques allow the analysis of many samples with relatively little effort. They have facilitated experiments aimed at using the concentrations of chemical species to explain sensory perceptions(*3;4*), determine the optimum duration and conditions of malaxation(*5-7*) and characterise olive cultivar and stage of ripeness(*8*). These techniques, though rapid compared with traditional reflux extraction techniques, still require at least an hour per sample. Both Tenax tubes and SPME fibres require at least 30 min exposure to the sample head space to collect an amount large enough for analysis. Quantitation must be made with care, as each adsorbent has different affinities for different VOCs, necessitating thorough calibration and continual use of an internal standard(*9*). As a consequence of the widespread use of these techniques, obtaining results can be straightforward due to the availability of autosamplers and specialised software. However, those results must still be interpreted.

### 3.1.2 Electronic Nose

Rapid, low-cost chemical analysis equipment such as the electronic nose has shown promise for olive oil VOC analysis in several studies(*10;11*). An electronic nose is a group of sensors, each of which is able to detect the presence of certain VOCs. Each sensor normally provides a single value. The

responses from all sensors in an electronic nose (assuming the sensors have been properly selected) form a unique pattern which can provide information on the sample being analysed. Electronic noses are gas sensors which can operate in ambient air (hence little sample preparation is necessary) and have short analysis times (usually a few minutes).

Electronic noses can be very complicated when it comes to interpreting their results, as multivariate techniques such as those described in appendix A are often needed. This is due to the low selectivity of most sensors used. The sensor responses are biased toward certain VOCs, yet they do not measure these exclusively. Therefore it is difficult to know which compounds an electronic nose is measuring without thorough investigation, while quantitative information on individual VOCs is all but impossible. However, pattern recognition methods are very powerful, and it is due to the deciphering power of multivariate methods that the use of electronic noses is so widespread for food and other applications involving VOC analysis(*12*).

## 3.2 Identified VOCs

As mentioned in section 1.2.2, the major route for VOC production in olives is the lipoxygenase (LOX) pathway. A brief overview has already been given. Figure 3.1 demonstrates the origins of some important volatile compounds from linolenic acid via enzyme reactions in the LOX pathway(*13*). There are many VOCs in olive oil which are not products of the LOX pathway – some are formed by oxidation, some by fermentation and other processes. Many processes form distinctive products enabling, to an extent, the history of the oil to be discovered through analysis of VOCs. The health of the olives, the age of

the oil or any prolonged contact with bacteria or fungi are all expressed through the VOC profile of an oil.

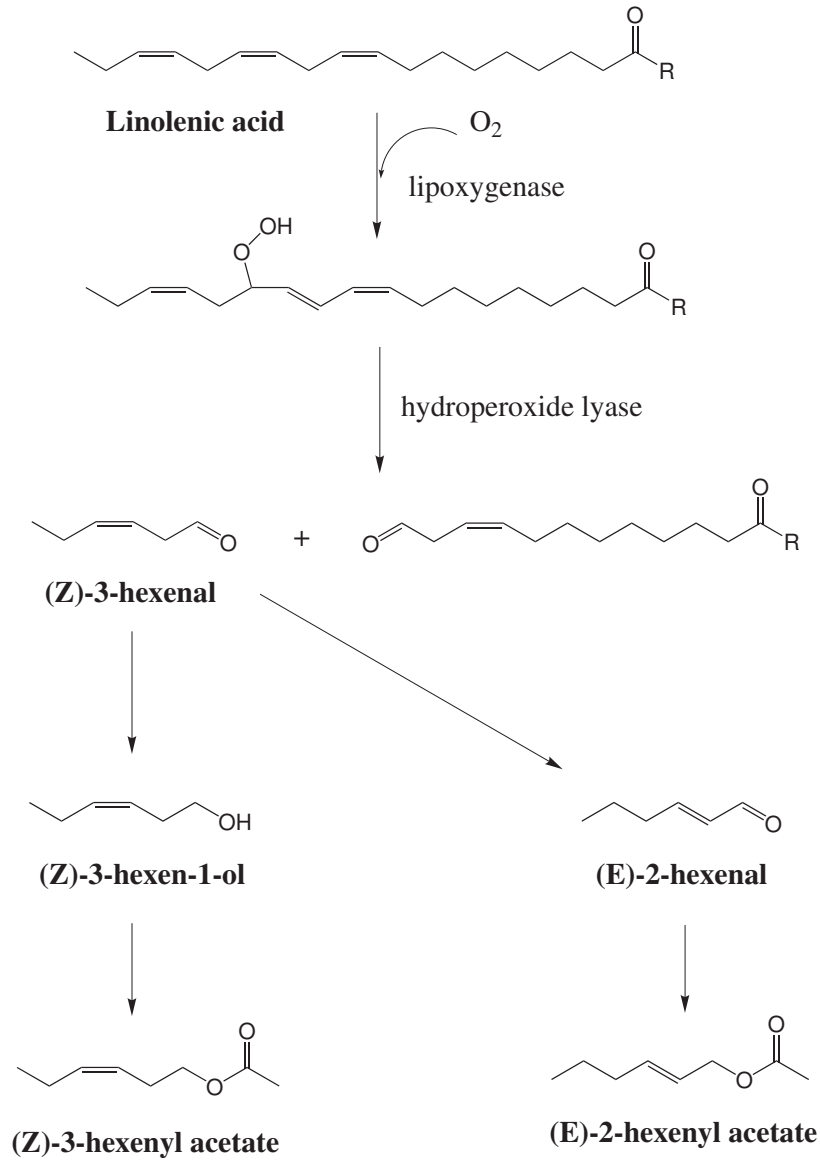



**Figure 3.1.** Volatile products derived from linolenic acid via the lipoxygenase pathway in olives. R may represent any of a wide selection of groups, although most commonly glycerol which is linked to the acid to form an ester and is also linked to two other fatty acids in the same way. All reactions are catalysed by enzymes, only the first two are listed here.

A list of important VOCs was obtained by Morales et. al.(*3*) by matching sensory and instrumental data using chemometrics. Of the 55 VOCs

observed, 33 were found to contribute significantly to the major sensory attributes considered in the study. Those identified are listed in table 3.1.

**Table 3.1.** Important VOCs identified by Morales et. al.(*3*), grouped by the sensory attributes they were found to be most highly correlated with.

| Attribute | Correlated compounds |
|---|---|
| green | methyl acetate, 1,3-hexadien-5-yne, 4-methyl pentan-2-one, 2-methyl-1-propanol, (Z)-3-hexenal, hexyl acetate, 3-hexenyl acetate, (Z)-2-penten-1-ol, (E)-2-hexen-1-ol, (Z)-3-hexen-1-ol |
| sweet | ethyl furan, ethyl propanoate, 1-penten-3-one, butyl acetate, hexanal |
| bitter and pungent | ethyl benzene, (E)-2-hexenal, (Z)-2-hexenal, 6-methyl-5-hepten-2-one, tridecene |
| undesirable | 1-penten-3-ol, 3-methyl butanol, 2-octanone, 1-hexanol, acetic acid |
| fruity | 2-butanone, 3-methyl butanal, 2-methyl butyl propanoate, ethenyl benzene, 2-nonanone |

## 3.3 Olive Oil Oxidation

The name given to the reaction of oxygen with lipids is 'autoxidation'(*14*). Ground-state triplet oxygen is too inert to oxidise lipids at an appreciable rate, and requires a catalyst to convert it to singlet oxygen (which is $10^3$-$10^4$ times more reactive than ground-state triplet oxygen at standard temperature(*15*)) before reaction can occur. Common catalysts are enzymes,

transition metals and pigments(*14*). Examples of these are lipoxygenase, copper and chlorophyll respectively. If the oil has been produced in clean conditions and the vegetable water adequately separated, only pigments should cause concern with regard to oil shelf-life. Pigments such as chlorophyll are photosensitisers, that is they require light to catalyse the triplet-to-singlet spin transition in oxygen.

Once formed, singlet oxygen adds across the double bonds of unsaturated lipids to form hydroperoxides(*16*). Hydroperoxides are primary lipid oxidation products. Hydroperoxides spontaneously decompose and initiate radical chain reactions. Many end products (secondary lipid oxidation products) are volatile and some have very low sensory detection thresholds(*17*). Hence, oxidation may seriously affect the sensory quality of olive oil, with only small changes (less than 2 %(*18*)) in the relative concentrations of fatty acids and triacylglycerols. The most obvious precautions to slow oxidation of olive oil are to minimise exposure to light and oxygen, store around 10-15 °C and ensure peroxide levels are low (this last condition cannot be directly controlled for virgin or extra virgin oils).

The susceptibility of an olive oil to oxidation depends on several factors. One factor is the fatty acid profile. Different fatty acids display different rates of oxidation; the rate of oxidation of oleic acid is 100 times that of stearic acid, that of linoleic acid is 1200 times, and linolenic acid oxidises 2500 times faster than stearic acid at 25 °C(*19*). Therefore, an oil with a higher linolenic acid concentration is likely to oxidise more rapidly than another oil under identical conditions. These fatty acids all undergo the same types of reactions during oxidation, however due to the differing number and positions of their double

bonds, each produces a different collection of products. There is the possibility of electron transfer between fatty acids, so the concentrations of oxidation products observed do not necessarily correspond to the concentrations of fatty acids present, even when their relative oxidation rates are taken into account(*20*).

### 3.3.1 The Rancimat Test

The Rancimat test is a widely used accelerated simulation of storage which determines an oil's resistance to oxidation. The resistance to oxidation can show the oxidative status of an oil at the beginning of the test. It is also sometimes used as an estimate of antioxidant capacity for oils(*21;22*), as oxidative stability has been shown to correlate with some measures of antioxidant capacity(*23*). This is true only in limited situations, and care must be taken when interpreting results from these studies. The antioxidant capacity of a sample is concerned only with the ability of certain components to scavenge radicals, whereas its resistance to oxidation relies on this and other factors.

The Rancimat test is carried out at approximately 120 °C (increased temperature is necessary for an accelerated test) and involves air bubbling through the oil sample. This air extracts any VOCs (oxidation products), depositing them in a sample of water. The conductivity of the water is constantly monitored – the point at which it reaches the maximum rate of change (the endpoint), is recorded. The time taken to reach the endpoint is called the 'induction time'(*15*).

This method is easy to use and generally reliable, however there are several drawbacks. The factors contributing to the induction time of an oil are not fully understood, making the analysis of results difficult in some

circumstances. At 120 °C the reactions occurring are different from those at standard temperature, with any compound not stable at this high temperature being lost. As a consequence, the contribution to the oxidative stability made by certain compounds may be incorrectly approximated, leading to an inaccurate result(*23*). Some authors have found that oil shelf-life is not strongly correlated with oxidative stability as determined by the Rancimat test(*24*), however it remains a convenient approximation.

## 3.4 Olive Oil Defects

There are several undesirable flavours or 'defects' which may appear in olive oil, each being produced under a certain set of conditions. Some important defects are listed below.

### 3.4.1 Rancid

This defect arises when oil is kept too long before consumption, it is the sensory result of the aforementioned oxidation processes. Rancidity is the most common defect encountered in olive oil, accordingly it is also the most commonly studied.

Morales, Rios and Aparicio(*25*) found the hexanal to nonanal peak area ratio in GC-MS analysis to be a reliable indicator of the progress of rancidity. This was due to the fact that hexanal is formed both by enzymatic oxidation in the lipoxygenase pathway and by non-biological oxidation, whereas nonanal is formed only through the latter process. Hence, the hexanal to nonanal ratio is high in fresh oil (in some cases up to 1900:1), and decreases towards 1:1 for rancid oil. Vichi et. al.(*26*) used Headspace Solid-Phase Microextraction (HS-SPME) coupled to GC-MS and GC-FID to investigate oxidation in olive oil, and arrived at a conclusion similar to that of Morales et. al. – nonanal is the most

reliable indicator of oxidative status. This was justified by stating that oleic acid – the fatty acid from which nonanal is produced – is not oxidised by enzymes, therefore any nonanal detected in olive oil must have been formed through non-biological (unfavourable) oxidation processes.

### 3.4.2 Winey

When olives are allowed to ferment, several different defects may be produced depending on the temperature and humidity. One such defect is responsible for 'winey' olive oil. Yeasts may ferment the olives when stored in a pile, producing ethanol and ethyl acetate. Bacteria of the genus *Acetobacter* may also metabolise the olives, producing acetic acid. These two processes often occur simultaneously, creating an aroma similar to wine and/or vinegar(*17*). Morales, Luna and Aparicio(*27*) presented a list of 16 representative volatile compounds correlated with this sensory attribute, most notably ethyl acetate, acetic acid, 1-pentanol and 2-butanol. Ethyl acetate was found to correlate most closely with the winey defect. Ethanol was also found to correlate with the winey defect, yet its correlation coefficient was fifteenth highest in the list of 16. It was still significant, however, at 0.91.

### 3.4.3 Fusty

The fusty defect is caused by much the same factors as the winey defect. The difference arises because the fermentation causing each defect is carried out by different bacteria. The practice of keeping olives in sacks during storage and transport to mills is largely responsible for fustiness. Bacteria of the genera *Enterobacteriaceae*, *Clostridia* and *Pseudomonas* are considered to cause the fusty defect(*17*). Angerosa et. al.(*28*) suggested the amount of 3-methyl-1-butanol present in olive oil was closely related to the strength of the fusty

defect. In their study, olives were stored in jute sacks with a capacity of 70 kg at 10 °C. Under these conditions, a sensory panel detected the fusty defect in the resultant oil after only 4 days of storage. The strength of the fusty defect was found to be expressed by equation 3.1, where F is the strength of the fusty defect as measured by a sensory panel, c is an empirical constant and the concentrations of the VOCs are measured by Gas-Liquid Chromatography (GLC), expressed in mg equivalents of the internal standard 1-nonanol per kg of oil.

$$F = c \frac{[3\text{-methyl-1-butanol}]}{\sqrt{[2\text{-hexenal}] + 1}} \quad (3.1)$$

### 3.4.4 Musty

The musty defect is another along the lines of winey and fusty. Musty olive oil is created from olives which have been stored improperly for an extended period of time and have been attacked by moulds. Musty olive oil displays significantly lower concentrations of C6 aldehydes and alcohols and higher concentrations of C8 compounds than good quality oil(*17*). The most prominent VOCs specific to musty olive oil are 1-octen-3-ol, (E)-2-heptenal, 1-octen-3-one and hexanal(*29*).

### 3.4.5 Muddy

Muddy olive oil is produced when sediments are allowed to reside in the oil. This situation generally occurs when the vegetable water has been only partly removed and/or the oil is unfiltered. A layer of sediment may accumulate on the bottom of the oil container and ferment. Muddy olive oil possesses an

acrid, acidic smell, due to butyric fermentation of the oil by bacteria on the surface of the sediment(*17*).

### 3.4.6 Frosted

When olives are exposed to temperatures below freezing, they may sustain significant damage due to intra- and extra-cellular frost formation. The damage includes burst (i.e. dead) cells and extreme dehydration, and severely affects the development of the fruit. The oil derived from such olives is much less bitter and pungent than good quality oil, and is also sweeter. It is generally not as resistant to oxidation, due to the break down of important antioxidant constituents(*30*). Frost damage is a large problem for many olive growers in New Zealand's South Island, with many regions receiving frosts before olives have properly ripened. Consequently, the timing of olive harvesting often revolves around the likelihood of frost events.

## 3.5 References


(1) Angerosa, F.; Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds in Virgin Olive Oil: Occurrence and Their Relationship With the Quality. *J. Chrom. A* **2004**, *1054*, 17-31.

(2) Flath, R. A.; Forrey, R. R.; Guadagni, D. G. Aroma Components of Olive Oil. *J. Agric. Food Chem.* **1973**, *21*, 948-952.

(3) Morales, M. T.; Alonso, M. V.; Ríos, J. J.; Aparicio, R. Virgin Olive Oil Aroma: Relationship between Volatile Compounds and Sensory Attributes by Chemometrics. *J. Agric. Food Chem.* **1995**, *43*, 2925-2931.

(4) Solinas, M.; Angerosa, F.; Cucurachi, A. Connessione tra Prodotti di Neoformazione Ossidativa delle Sostanze Grasse e Insorgenza del Difetto di Rancidità all'Esame Organolettico. Nota 1. *Riv. Soc. It. Sci. Alim.* **1985**, *14*, 361-368.

(5) Angerosa, F.; Mostallino, R.; Basti, C.; Vito, R. Influence of Malaxation Temperature and Time on the Quality of Virgin Olive Oils. *Food Chem.* **2001**, *72*, 19-28.

(6) Kalua, C. M.; Bedgood, D. R. J.; Bishop, A. G.; Prenzler, P. D. Changes in Volatile and Phenolic Compounds with Malaxation Time and Temperature during Virgin Olive Oil Production. *J. Agric. Food Chem.* **2006**, *54*, 7641-7651.

(7) Servili, M.; Selvaggini, R.; Taticchi, A.; Esposto, S.; Montedoro, G. Volatile Compounds and Phenolic Composition of Virgin Olive Oil: Optimization of Temperature and Time of Exposure of Olive Pastes to Air Contact during the Mechanical Extraction Process. *J. Agric. Food Chem.* **2003**, *51*, 7980-7988.

(8) Kalua, C. M.; Mailer, R. J.; Ayton, J.; Allen, M. S.; Bedgood, D. R. J.; Bishop, A. G.; Prenzler, P. D. Discrimination of Olive Oils and Fruits into Cultivars and Maturity Stages Based on Phenolic and Volatile Compounds. *J. Agric. Food Chem.* **2005**, *53*, 8054-8062.

(9) Contini, M.; Esti, M. Effect of the Matrix Volatile Composition in the Headspace Solid-Phase Microextraction Analysis of Extra Virgin Olive Oil. *Food Chem.* **2006**, *94*, 143-150.

(10) Aparicio, R.; Rocha, S. M.; Delgadillo, I.; Morales, M. T. Detection of Rancid Defect in Virgin Olive Oil by the Electronic Nose. *J. Agric. Food Chem.* **2000**, *48*, 853-860.

(11) García-Gonzalez, D. L.; Aparicio, R. Detection of Vinegary Defect in Virgin Olive Oils by Metal Oxide Sensors. *J. Agric. Food Chem.* **2002**, *50*, 1809-1814.

(12) Di Natale, C.; Macagnano, A.; Paolesse, R.; D'Amico, A. Artificial Olfaction Systems: Principles and Applications to Food Analysis. *Biotechnol. Agron. Soc. Environ.* **2001**, *5*, 159-165.

(13) Sánchez, J.; Harwood, J. L. Biosynthesis of Triacylglycerols and Volatiles in Olives. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 564-573.

(14) Boskou, D. *Olive Oil: Chemistry and Technology*; AOCS Press: Champaign, IL, USA, 1996.

(15) Velasco, J.; Dobarganes, C. Oxidative Stability of Virgin Olive Oil. *Eur. Food Res. Technol.* **2002**, *104*, 661-676.

(16) Foote, C. S. Photosensitized Oxygenations and the Role of Singlet Oxygen. *Acc. Chem. Res.* **1968**, *1*, 104-110.
(17) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 639-660.
(18) Gomez-Alonso, S.; Salvador, M.; Fregapane, G. Evolution of the Oxidation Process in Olive Oil Triacylglycerol under Accelerated Storage Conditions (40-60°C). *J. Amer. Oil Chem. Soc.* **2004**, *81*, 177-184.
(19) Belitz, H.-D.; Grosch, W. *Food Chemistry*, 2nd English ed.; Springer-Verlag: Berlin, Germany, 1999.
(20) Anonymous, Private communication, response to submitted journal article, 2007.
(21) Hudson, B.; Lewis, J. Polyhydroxy Flavonoid Antioxidants for Edible Oils. Structural Criteria for Activity. *Food Chem.* **1983**, *10*, 47-55.
(22) Dziedzic, S.; Hudson, B. Phenolic Acids and Related Compounds as Antioxidants for Edible Oils. *Food Chem.* **1984**, *14*, 45-51.
(23) Ninfali, P.; Bacchiocca, M.; Biagiotti, E.; Servili, M.; Montedoro, G. Validation of the Oxygen Radical Absorbance Capacity (ORAC) Parameter as a New Index of Quality and Stability of Virgin Olive Oil. *J. Amer. Oil Chem. Soc.* **2002**, *79*, 977-982.
(24) Dijkstra, A.; Maes, P.; Meert, D.; Meeussen, W. *Proceedings of the World Congress of the International Society for Fat Research, The Hague, Netherlands*; Oils-Fats-Lipids; Vol. 3, p 629-637.
(25) Morales, M.; Rios, J.; Aparicio, R. Changes in the Volatile Composition of Virgin Olive Oil during Oxidation: Flavors and Off-Flavors. *J. Agric. Food Chem.* **1997**, *45*, 2666-2673.
(26) Vichi, S.; Pizzale, L.; Conte, L. S.; Buxaderas, S.; López-Tamames, E. Solid-Phase Microextraction in the Analysis of Virgin Olive Oil Volatile Fraction: Modifications Induced by Oxidation and Suitable Markers of Oxidative Status. *J. Agric. Food Chem.* **2003**, *51*, 6564-6571.
(27) Morales, M. T.; Luna, G.; Aparicio, R. Sensory and Chemical Evaluation of Winey-Vinegary Defect in Virgin Olive Oils. *Eur. Food Res. Technol.* **2000**, *211*, 222-228.
(28) Angerosa, F.; Di Giacinto, L.; Solinas, M. Influenza dello Stoccaggio in Massa delle Olive sull'Aroma degli Oli di Risulta: Valutazione del Difetto di "Riscaldo" Mediante Analisi HPLC e GLC dei Componenti Volatili. *Riv. Merceol.* **1990**, *29*, 275-294.
(29) Morales, M.; Luna, G.; Aparicio, R. Comparative Study of Virgin Olive Oil Sensory Defects. *Food Chem.* **2005**, *91*, 293-301.
(30) Morelló, J.-R.; Motilva, M.-J.; Ramo, T.; Romero, M.-P. Effect of Freeze Injuries in Olive Fruit on Virgin Olive Oil Composition. *Food Chem.* **2003**, *81*, 547-553.

# Chapter 4

## VOC Analysis Method

### 4.1 Introduction

The starting point for SIFT-MS analysis of any sample is assignment of the observed product masses to the compounds. Mass scans are used for this purpose. This is a task where the difficulty varies with the number of components in the mixture. Once the list of sample components has been decided on, a method is generated whereby SIM scans are used to measure the appropriate product masses and provide the real-time concentration data that characterise SIFT-MS as an analytical technique. The path from first analysis of olive oil head space to the development of a SIM method able to be used for quantitation is outlined in this chapter.

### 4.2 Initial Investigation

Mass scans of head space samples were performed on a selection of olive oils using both $H_3O^+$ and $NO^+$, the chosen reagent ions. It became apparent at an early stage that $O_2^+$ was less suitable for olive oil analysis, or indeed the analysis of any complicated mixture, due to the cluttered nature of its spectra. Reactions of $O_2^+$ with many analyte molecules induce molecular fragmentation, resulting in multiple products for almost all analytes. When analysing a system composed of more than ten or fifteen analytes using $O_2^+$, the possibility of product ions from different analytes being mass coincident is significant and hence it is difficult to identify – let alone quantify – the analytes. Ion-molecule reactions involving $H_3O^+$ and $NO^+$ have

fewer reaction pathways and therefore these two reagent ions were deemed sufficient to provide analyte identification. However, coincident masses may still be observed for different analytes using these reagent ions, meaning selection of the most appropriate products to measure from each analyte is done on a case-by-case basis.

33 samples of New Zealand olive oils from the 2002 annual New Zealand Olive Association (now Olives New Zealand) Extra Virgin Olive Oil Awards were obtained. 10 mL of each oil was placed in a 125 mL glass vial and sealed with a silicon rubber septum. The vials were stored under nitrogen gas in a cool, dark cupboard, with equilibrium of the volatile components between the liquid and the head space well achieved before analysis. The head space of each vial was sampled through a stainless steel capillary by SIFT-MS, while the oil was purged with nitrogen at the same flow rate of approximately 90 mL $min^{-1}$. All vials were sampled in this way generating mass scans by chemical ionisation with the $H_3O^+$ and $NO^+$ reagent ions. Product ions between $m/z = 10$ and $m/z = 155$ were measured. Each scan took approximately 17 seconds. Typical spectra from these oils are shown in figures 4.1 and 4.2.

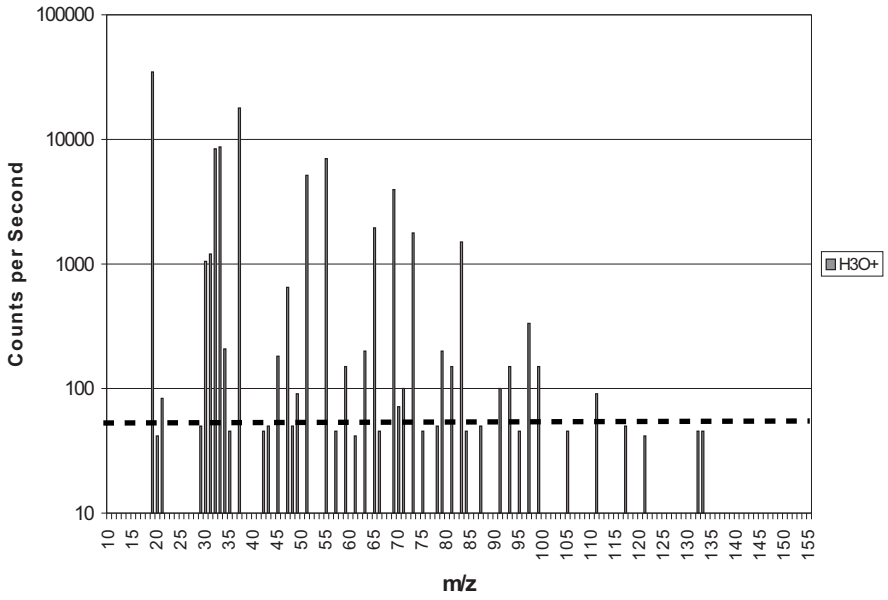



**Figure 4.1.** A mass scan with the $H_3O^+$ reagent ion of the head space above a sample of olive oil. The acquisition time for this scan was just over 17 seconds. The scale on the y axis is logarithmic. Note the randomly distributed noise of 50 counts per second, which corresponds to a single count within the time period of the measurement (20 ms for each mass value). All mass scans in this thesis include a dotted line, below which any peaks are considered to be randomly placed noise picked up by the sensitive electronics used in SIFT-MS instruments. Peaks below the dotted lines do not correspond to ions and should be ignored.

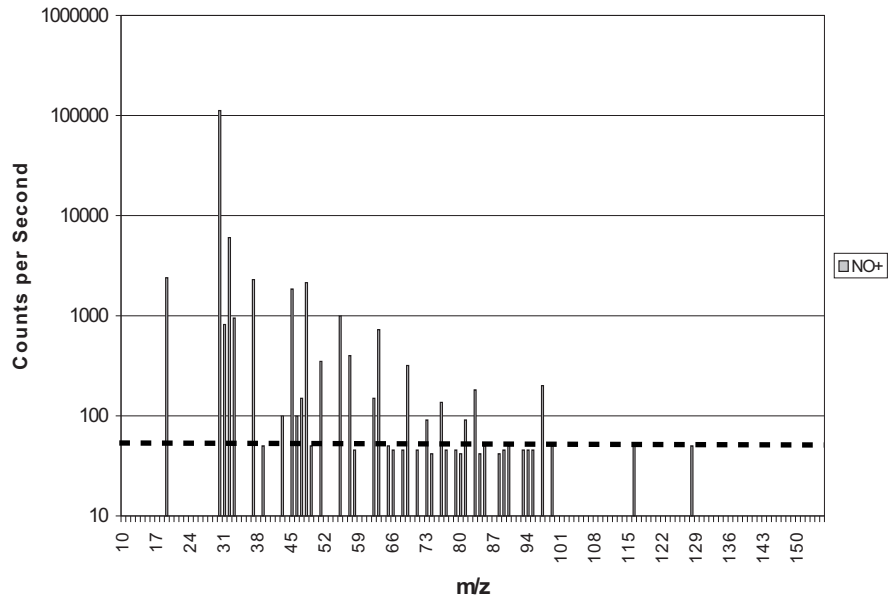



**Figure 4.2.** A mass scan with the $NO^+$ reagent ion of the head space above a sample of olive oil. This spectrum has fewer peaks than does figure 4.1, primarily because the ions involved in $NO^+$ reactions are less likely to form water clusters than those involved in $H_3O^+$ reactions.

From analysis of the award-winning olive oils, a number of product masses were identified. The m/z values representing products of SIFT-MS reactions to be assigned to olive oil VOCs are listed in table 4.1. The remainder of this chapter deals with the assignment of these products to analyte compounds.

**Table 4.1.** Product m/z values observed during analysis of 33 extra virgin olive oils. The reagent ions used to obtain the products are shown. Not all product masses listed were observed in all oils.

| $H_3O^+$ (m/z = 19) | | | $NO^+$ (m/z = 30) | | |
|---|---|---|---|---|---|
| 31 | 65 | 95 | 45 | 87 | 119 |
| 33 | 69 | 97 | 57 | 88 | 128 |
| 43 | 71 | 99 | 59 | 90 | 141 |
| 45 | 77 | 101 | 63 | 97 | 150 |
| 47 | 79 | 111 | 69 | 99 | |
| 49 | 81 | 119 | 70 | 104 | |
| 51 | 83 | 125 | 71 | 108 | |
| 59 | 85 | 139 | 77 | 114 | |
| 61 | 89 | 143 | 81 | 115 | |
| 63 | 93 | | 83 | 118 | |

## 4.3 Assignment of Product Masses

### 4.3.1 Ethanol and Methanol

The most prominent features of the majority of the $H_3O^+$ spectra were the peaks at m/z = 33, 47, 51, 65, 69 and 83. All of these peaks may be confidently assigned to protonated methanol and ethanol ions and clusters of these ions with water(*1*). Reactions representative of those which form the aforementioned products are shown in reactions 4.1-3.

$$H_3O^+ + CH_3OH \rightarrow CH_3OH.H^+ (m/z = 33) + H_2O \quad (4.1)$$

$$H_3O^+ + C_2H_5OH \rightarrow C_2H_5OH.H^+ (m/z = 47) + H_2O \quad (4.2)$$

$$CH_3OH.H^+ + H_2O + M \rightarrow CH_3OH.H^+.H_2O (m/z = 51) + M \quad (4.3)$$

Reaction 4.3 is a clustering reaction where the initially formed cluster ion at m/z = 51 requires a third body (M, normally a helium atom) to accept the excess energy brought into the cluster ion by the reactants and stabilise it. If the third body is not encountered in time, the reaction quickly proceeds in the reverse direction and no cluster is observed.

The association of ethanol with water in the SIFT-MS flow tube is very similar to that of methanol (reaction 4.3), and results in a cluster ion at m/z = 65 (the parent protonated ion from reaction 4.2 at m/z = 47 plus 18 amu introduced by water). These clusters often attract one more water molecule each to form species with m/z = 69 and 83 containing methanol and ethanol respectively. Only clusters comprising three species (an alcohol and two water molecules in these cases) are able to survive the flow tube intact, therefore no larger water clusters are observed for these species. There are, however, dimerised species which may further complicate the spectra. Molecular methanol and ethanol may be present at such concentrations (typically greater than 20 ppm/mol in a sample when a sample flow of 1-2 Torr L $s^{-1}$ is used) that instead of water associating with the protonated species as in reaction 4.3, another methanol or even ethanol molecule may take its place. This leads to large peaks at m/z = 65 (reaction 4.4) and 83 (reaction 4.5) when methanol is present to excess (a methanol dimer and its associated single water cluster), m/z = 93 and 111 when ethanol is abundant, and additionally 79 and 97 when both methanol and ethanol are present at high concentrations.

$$CH_3OH.H^+ + CH_3OH + M \rightarrow (CH_3OH)_2.H^+ (m/z = 65) + M \qquad (4.4)$$

$$(CH_3OH)_2.H^+ + H_2O + M \rightarrow (CH_3OH)_2.H^+.H_2O\ (m/z = 83) + M \quad (4.5)$$

It is important to measure all clusters of a compound in order to obtain an accurate estimate of the concentration in the sample, as a change in humidity or in the concentrations of other components may induce the conversion of different proportions of primary products into secondary products. This means that the contribution of secondary products (clusters) is not predictable (unlike the branching ratios of primary products) and must always be measured. Therefore, the fact that the ethanol-water clusters at m/z = 65 and 83 coincide with the methanol dimer found at m/z = 65 and its associated water cluster at m/z = 83 is potentially a difficult hurdle to overcome in the quantification of these two species when both are present above several ppm in a sample.

The data obtained from the 33 olive oil samples were used to investigate this situation further. The accumulation of a water molecule by the m/z = 47 ion to form the m/z = 65 cluster should depend directly on how much of the m/z = 47 ion is formed (which itself depends on the concentrations of $H_3O^+$ and ethanol in the flow tube), as all were carried out in an air-conditioned environment at approximately the same humidity. This would produce a direct linear relationship between the number of counts of m/z = 47 and m/z = 65, with the slope of the line dependent on the flow tube humidity and the clustering rate of the protonated ethanol ion with water. The occurrence of the methanol dimer will follow similar

lines, with the relationship between the monomer (at m/z = 33) and the dimer (at m/z = 65) being dependent on the concentration of methanol in the flow tube.

As can be seen in figure 4.3, for a selection of 40 olive oils analysed (the 33 mentioned before and seven other olive oils analysed on the same day), where ethanol and methanol were both present at appreciable concentrations, there are very good correlations between the intensities of the peaks at m/z = 47 and 65 and between those at m/z = 33 and 51, an acceptable correlation between m/z = 47 and 93 (m/z = 93 is the $(C_2H_5OH)_2.H^+$ ion) and a very poor correlation between m/z = 33 and 65 (m/z = 65 can be either the $C_2H_5OH.H^+.H_2O$ or $(CH_3OH)_2.H^+$ ions). This shows that the intensities of both water cluster and dimer products are closely related to those of the corresponding primary products with the exception of the methanol dimer peak at m/z = 65. As the intensity of the m/z = 65 peak shows good correlation with that of the m/z = 47 peak and very poor correlation with that of the m/z = 33 peak, this suggests that by far the bulk of the m/z = 65 peak is originating from the m/z = 47 ion. This is reinforced by the fact that the m/z = 33 peak reaches more than three times the intensity of the m/z = 47 peak and therefore would display a much better correlation with the m/z = 65 peak if both were forming this product at an equal rate. It is worth noting that the data suggest a non-linear relationship between the intensities of the m/z = 47 ion and its dimer at m/z = 93. A non-linear relationship may well exist, as the rate of dimer formation is expected to increase as the ethanol concentration increases. However, the low number of points in the appropriate region of the chart and the poor agreement between them prohibit the characterisation of the relationship. The data in

figure 4.3 add weight to the argument that methanol does not form significant amounts of its dimer product under SIFT-MS conditions, although the only way to be sure is to measure the intensity of the m/z = 65 peak generated from a sample which contains methanol and no ethanol. Figure 4.4 contains a SIM scan of an olive oil displaying as close to that situation as possible.

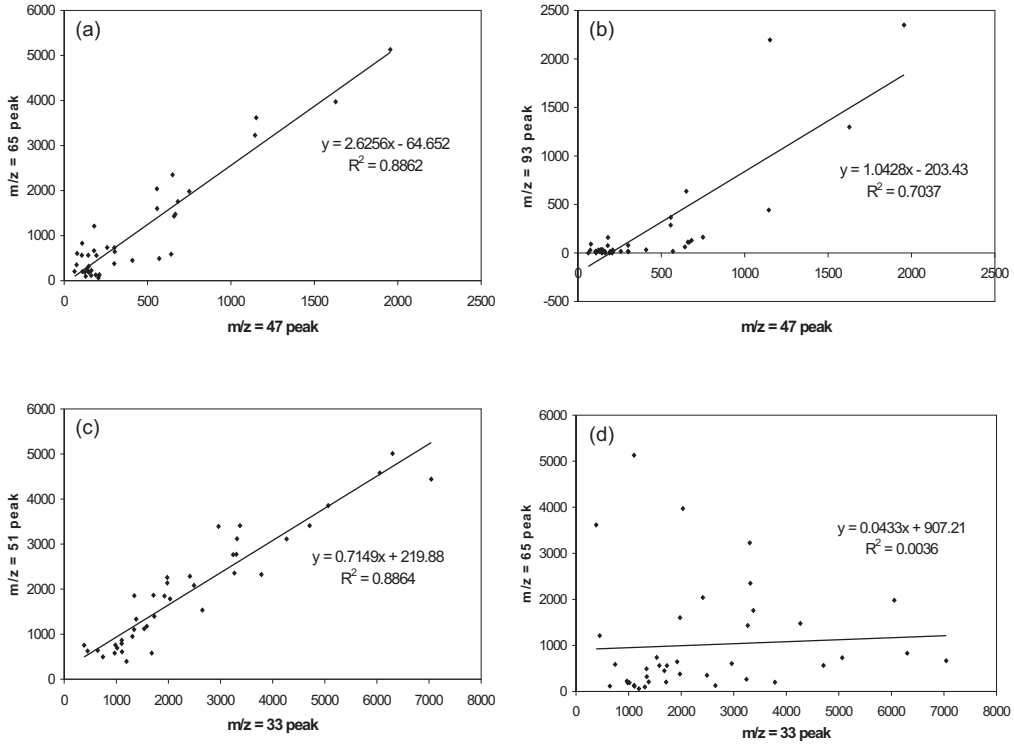



**Figure 4.3.** Correlation between intensities of the peaks at **(a)** m/z = 47 and 65, the primary ethanol product and ethanol water cluster respectively, **(b)** m/z = 47 and 93, the primary ethanol product and ethanol dimer cluster respectively, **(c)** m/z = 33 and 51, the primary methanol product and methanol water cluster respectively and **(d)** m/z = 33 and 65, the primary methanol product and possibly methanol dimer cluster respectively from reaction with the $H_3O^+$ reagent ion for 40 olive oil samples.

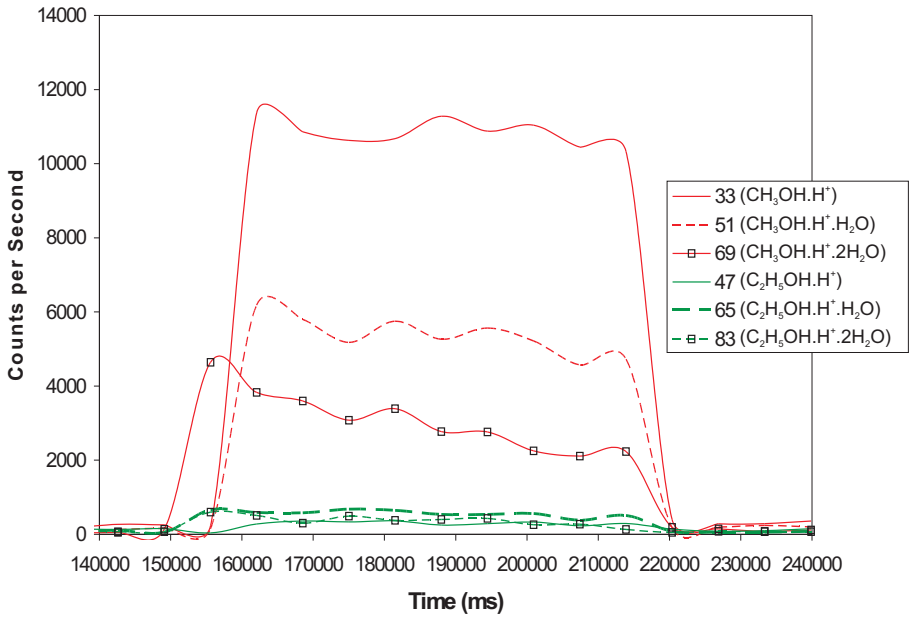



**Figure 4.4.** A SIM scan of a refined olive oil with 5 ppm/vol methanol added. The peaks shown are all derived from $H_3O^+$ and all correspond to methanol or ethanol. Ethanol was observed in this sample (no olive oil has been encountered which did not contain at least a low concentration of ethanol), however its concentration was much lower than that of methanol. It can be seen that the intensity of the m/z = 65 peak is at the level expected by formation of a water cluster from the product at m/z = 47 and is not contributed to significantly by any methanol products.

The primary peak observed in the $NO^+$ spectrum for ethanol was at m/z = 45, corresponding to loss of $H^-$ as shown in reaction 4.6(*1*). The water clusters of this ion (similar to those shown in reactions 4.4 and 4.5 and mentioned in the accompanying text) were also observed at m/z = 63 and 81. The primary peak observed in the $NO^+$ spectrum with methanol was due to association (as in reaction 4.7(*1*)) and found at m/z = 62, but it was not always observed as the rate

coefficient for the reaction of $NO^+$ with methanol is less than one hundredth of that with ethanol(*1*). Hence, the intensity of the m/z = 62 peak for methanol was sometimes below the detection limit of the method used.

rate constant = 1.2 x $10^{-9}$ $cm^3s^{-1}$

$$NO^+ + C_2H_5OH \rightarrow C_2H_5O^+ (m/z = 45) + HNO \quad (4.6)$$

rate constant = 1 x $10^{-11}$ $cm^3s^{-1}$

$$NO^+ + CH_3OH + M \rightarrow CH_3OH.NO^+ (m/z = 62) + M \quad (4.7)$$

Once the mass peaks in each spectrum were assigned to ethanol and methanol and it was accepted that correct qualitative information was available from SIM scans for these two analytes, confirmation of correct quantitative information came from measurements of the Henry's law coefficients for each analyte. HPLC grade ethanol and methanol were obtained and mixed with water so as to make a number of solutions with varying concentrations. 5 mL of each mixture was added to a 500 mL Schott bottle and left for 20 min at room temperature. The head space was then analysed for ethanol and methanol by SIFT-MS, using the masses discussed above for each compound, and rate constants taken from the literature(*1*). Literature values for Henry's law coefficients were obtained from Sander(*2*), given in mol $L^{-1}$ $atm^{-1}$. The range of literature values for methanol was 140-230 with a mean of 203 mol $L^{-1}$ $atm^{-1}$, while that for ethanol was 120-220 with a mean of 184 mol $L^{-1}$ $atm^{-1}$. As can be seen in figure 4.5, there is agreement between the Henry's law coefficients obtained by other methods and those

determined by SIFT-MS (the slopes of the curves shown, $k_H = 160 \pm 10$ mol $L^{-1}$ $atm^{-1}$ and $k_H = 150 \pm 10$ mol $L^{-1}$ $atm^{-1}$ for ethanol and methanol respectively) for these two VOCs. This is reassuring, as it shows that the accuracy of SIFT-MS measurements of ethanol and methanol throughout this research is acceptable. It should also be mentioned that the SIFT-MS measurements are often made at significantly lower concentrations than many of the more traditional methods for evaluating Henry's law coefficients.

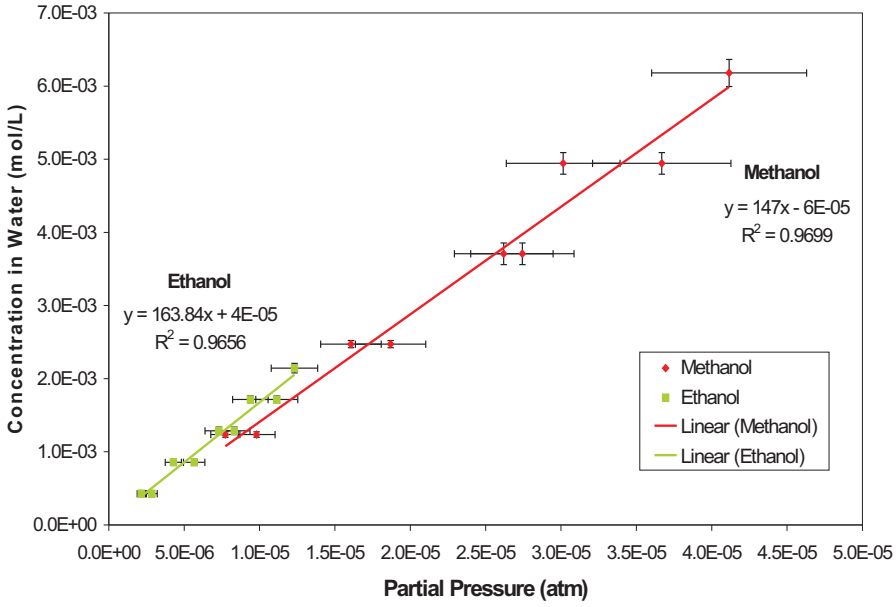



**Figure 4.5.** Henry's Law coefficients determined for methanol and ethanol from aqueous solutions at 25 °C are shown as the slopes of the corresponding curves.

### 4.3.2 Acetone and Propanal

The characteristic peaks for acetone were also observed in the $H_3O^+$ spectra of these oils. With $H_3O^+$ as the source of chemical ionisation, acetone displays peaks at m/z = 59, 77 and 95 – these are explained in reactions 4.8, 4.9 and 4.10(*3*),

where M is the same as in reaction 4.3. A rapid reaction of $NO^+$ with acetone gives the association at m/z = 88 (reaction 4.11). However, an isomer of acetone was also identified in the $NO^+$ spectrum, giving a product at m/z = 57. Due to the different reaction products of aldehydes and ketones with $NO^+$ (reactions 4.11 and 4.12, association and hydride abstraction respectively)(*3*), propanal was also identified in olive oil head space.

$$H_3O^+ + CH_3COCH_3 \rightarrow CH_3COCH_3.H^+ (m/z = 59) + H_2O \quad (4.8)$$

$$CH_3COCH_3.H^+ + H_2O + M \rightarrow CH_3COCH_3.H^+.H_2O (m/z = 77) + M \quad (4.9)$$

$$CH_3COCH_3.H^+.H_2O + H_2O + M \rightarrow CH_3COCH_3.H^+.2H_2O (m/z = 95) + M \quad (4.10)$$

$$NO^+ + CH_3COCH_3 + M \rightarrow CH_3COCH_3.NO^+ (m/z = 88) + M \quad (4.11)$$

$$NO^+ + C_2H_5CHO \rightarrow C_2H_5CO^+ (m/z = 57) + HNO \quad (4.12)$$

As propanal and acetone are both present in olive oil and both react in the same way with $H_3O^+$, only the sum of the acetone and propanal concentrations is provided by the $H_3O^+$ reagent ion. $NO^+$ alone was considered sufficient to measure these two analytes as it allows each analyte to be quantified individually.

Acetone and propanal were also the subjects of a Henry's law investigation (figure 4.6). Propanal in particular is a product of olive oil oxidation, making accurate quantitative analysis important. The same method was used for these compounds as for ethanol and methanol, with the results of the Henry's law measurements again being compared with data from Sander(*2*). The range for acetone from Sander was 3-35 with a mean of 24 mol $L^{-1}$ $atm^{-1}$, while that for propanal was 2.8-13 with a mean of 9 mol $L^{-1}$ $atm^{-1}$. The value obtained for acetone ($20 \pm 2$ mol $L^{-1}$ $atm^{-1}$) was in good agreement with the literature, however that for propanal ($18 \pm 2$ mol $L^{-1}$ $atm^{-1}$) was not. The difference was considered minor when compared with the five-fold variation in reported values, therefore subsequent propanal concentrations were not altered on the basis of this experiment.

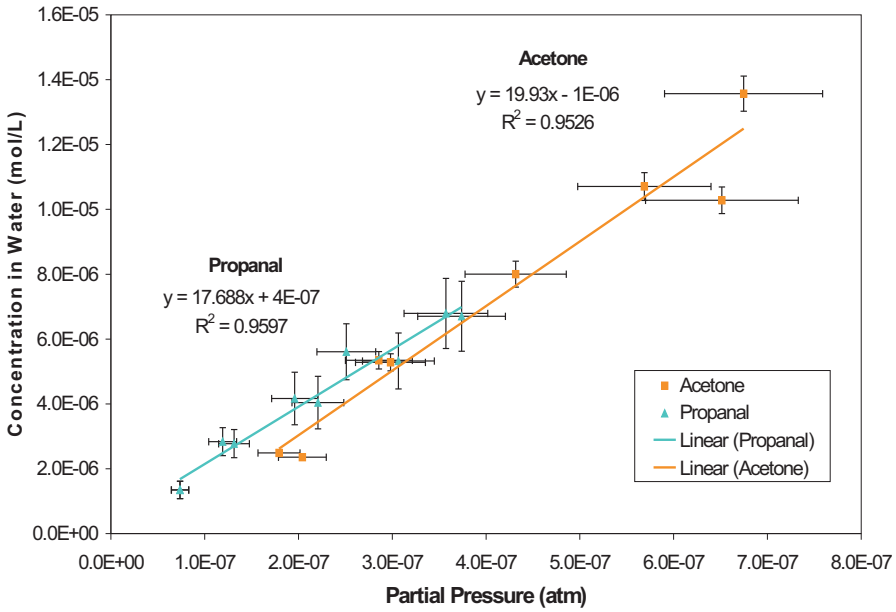
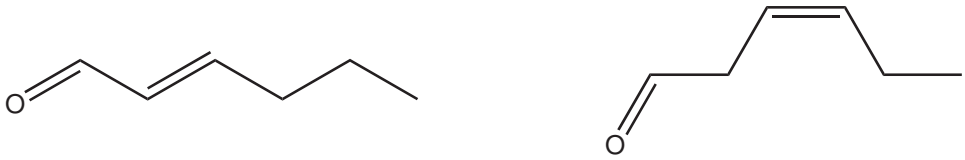



**Figure 4.6.** Henry's Law coefficients determined for acetone and propanal from aqueous solutions at 25 °C are shown as the slopes of the corresponding curves.

### 4.3.3 Hexenal Isomers

The most abundant VOC in the head space above olive oil according to many researchers is (E)-2-hexenal(*4-7*). Although SIFT-MS cannot reliably distinguish between isomers of some analytes such as different xylenes, mesitylenes(*8*) or terpenes(*9*), it is able to distinguish between (E)-2-hexenal and (Z)-3-hexenal(*3*). Both of these unsaturated aldehydes have been found by other researchers in olive oil head space(*4-7*), and while other isomers may also be present, the assignment of (E)-2-hexenal is considered to be reliable. The products obtained from the reaction of (E)-2-hexenal under SIFT-MS conditions have been reported as m/z = 99 with the $H_3O^+$ reagent ion, and m/z = 71 and 97 with $NO^+$(*3*) (reactions 4.13, 4.14 and 4.15). Due to coincident masses with products of other VOCs, only m/z = 97 and 115 (the water cluster of the m/z = 97 ion) from the $NO^+$ reagent ion were measured for this analyte. As there are two competing reactions in this case and the m/z = 97 ion comprises 85 % of the ionic products, its intensity was multiplied by 1.18 so that the m/z = 71 product (which coincides with the single primary product of butanal(*3*)) need not be measured. A standard of pure (E)-2-hexenal, obtained from Bedoukian Research Ltd. (Danbury, CT, USA), was used to ensure the accuracy of the m/z values being monitored. There is no way to be sure if other isomers contribute to the signal assigned to (E)-2-hexenal (because standards of all possible isomers are not available), yet the evidence supplied by other researchers leads to the assumption that the contributions of other hexenal isomers are minor.



(E)-2-hexenal (Z)-3-hexenal

**Figure 4.7.** Structures of (E)-2-hexenal and (Z)-3-hexenal, two hexenal isomers found in olive oil headspace. Both are produced via the lipoxygenase pathway.

$$H_3O^+ + C_3H_7CHCHCHO \rightarrow C_3H_7CHCHCHO.H^+ (m/z = 99) + H_2O \quad (4.13)$$

$$NO^+ + C_3H_7CHCHCHO \rightarrow C_3H_7CHCHCO^+ (m/z = 97, 85\ \%) + HNO \quad (4.14)$$

$$NO^+ + C_3H_7CHCHCHO \rightarrow C_4H_7O^+ (m/z = 71, 15\ \%) + CH_3NCO \quad (4.15)$$

The accuracy of (E)-2-hexenal quantification was investigated by a Henry's law study. The results of this study are shown in figure 4.8. Only one value was available for the Henry's law coefficient of (E)-2-hexenal from other researchers(*2*). The literature value of 20 mol $L^{-1}$ $atm^{-1}$ is close to that of 15 ± 1 mol $L^{-1}$ $atm^{-1}$ found in the present study. While these values do not agree, the wide variation in literature values observed for the other compounds in the study would be expected to be repeated here had there been more than one literature value for (E)-2-hexenal. The SIFT-MS quantification of (E)-2-hexenal was considered acceptable.

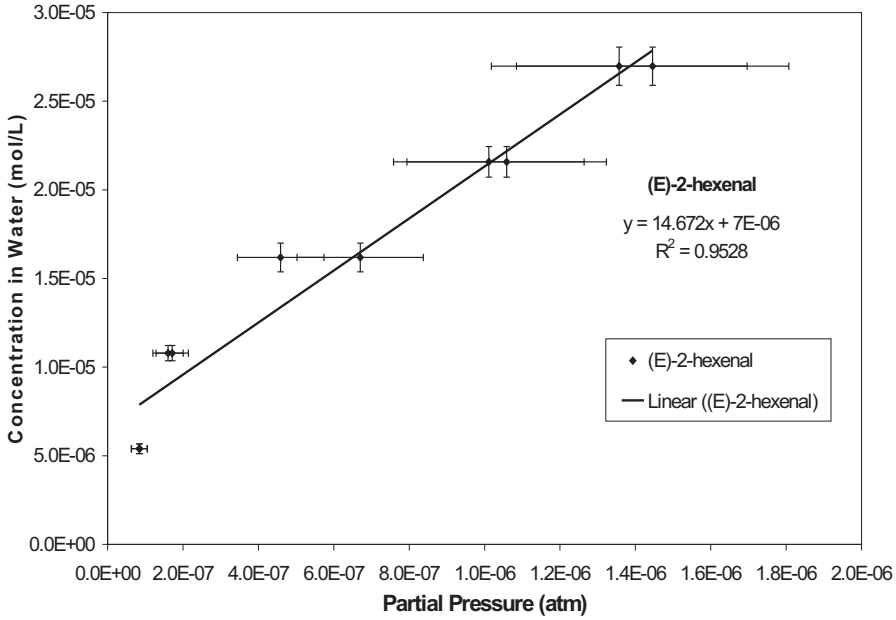



**Figure 4.8.** Henry's Law coefficient determined for (E)-2-hexenal from aqueous solution at 25 °C is shown as the slope of the curve.

(Z)-3-hexenal forms ions at m/z = 81 and 99 with the $H_3O^+$ reagent ion (reactions 4.16 and 4.17) and m/z = 69, 70 and 97 with $NO^+$ (reactions 4.18, 4.19 and 4.20)(*3*). Several of these m/z values coincide with those of other analytes, most notably (E)-2-hexenal. To distinguish the two analytes, it was assumed that (E)-2-hexenal would always be present at significantly higher concentration than (Z)-3-hexenal, therefore any product peaks which coincide for these two analytes would contain only minor contributions from (Z)-3-hexenal and could be assumed to be solely due to (E)-2-hexenal. Both of these analytes are products of the lipoxygenase pathway, and (E)-2-hexenal is by far the major product, so this is almost certainly a valid assumption. As the (Z)-3-hexenal m/z = 81 peak with the $H_3O^+$ reagent ion coincides with a water cluster ion for acetaldehyde, $NO^+$ was

considered the only useful reagent ion for measuring (Z)-3-hexenal. Two products were measured for (Z)-3-hexenal: m/z = 69 and 70. These ions make up 30 % and 20 % of the primary products respectively. To obtain an accurate estimation of the analyte concentration, the intensity of each signal was multiplied by two and summed.

$$H_3O^+ + C_2H_5CHCHCH_2CHO \rightarrow C_6H_9^+ (m/z = 81, 65\ \%) + 2H_2O \quad (4.16)$$

$$H_3O^+ + C_2H_5CHCHCH_2CHO \rightarrow C_6H_{11}O^+ (m/z = 99, 35\ \%) + H_2O \quad (4.17)$$

$$NO^+ + C_2H_5CHCHCH_2CHO \rightarrow C_6H_9O^+ (m/z = 97, 50\ \%) + HNO \quad (4.18)$$

$$NO^+ + C_2H_5CHCHCH_2CHO \rightarrow C_5H_9^+ (m/z = 69, 30\ \%) + HCNO_2 \quad (4.19)$$

$$NO^+ + C_2H_5CHCHCH_2CHO \rightarrow C_4H_6O^+ (m/z = 70, 20\ \%) + C_2H_4NO \quad (4.20)$$

The qualitative assignment of m/z values to (Z)-3-hexenal was also shown to be correct through the use of a standard, however the quantitative assignment is less certain, due to concerns raised by the low nasal detection threshold of (Z)-3-hexenal in humans(*10*). The accuracy of the concentration calculation for this analyte was not investigated by a Henry's Law experiment. The Henry's Law constant for (Z)-3-hexenal was not mentioned in Sander(*2*), which is considered to

be the most authoritative list of Henry's Law constants yet published, and could not be obtained from any other source. With no value available for the expected Henry's Law constant, little would be gained from the corresponding experiment. One complication to the method devised for (Z)-3-hexenal is that other VOCs may contribute to the signals at m/z = 69 and 70. For example, in some olive oil samples (Z)-3-hexenal was measured to have quite high concentrations which, if true, would be expected to cause the aroma of the oil to be substantially dominated by this one compound. For example, an average concentration calculated for (Z)-3-hexenal from the selected masses is 230 ppb. The nasal detection threshold of (Z)-3-hexenal for humans is 1.7 ppb(*10*), over 130 times lower. At high concentrations this compound has a sharp grassy, almost unpleasant aroma, and no olive oil yet encountered in this work has been found to possess such an aroma.

Due to the complicated nature of the spectra and the high number of compounds which could potentially contribute to the peaks at m/z = 69 and 70, it is not known to what extent the estimates of (Z)-3-hexenal are affected in some oils. Of course, the interaction of aroma compounds with the human senses is not fully understood. It might also be the case that the (Z)-3-hexenal concentrations may be accurate, and there may be interactions with other compounds such as methanol and ethanol during sensory analysis which can mask the dominance of this compound and produce a more balanced aroma(*11*). It was decided to monitor the masses associated with (Z)-3-hexenal but treat them with some scepticism, in case they are overestimating the (Z)-3-hexenal presence.

### 4.3.4 Propanol

Reaction of propanol with $H_3O^+$ produces m/z = 43 and 61 (reactions 4.21 and 4.22), and with $NO^+$ 59, 77 and 119 (reactions 4.23, 4.24 and 4.25) are formed(*1*). Acetic acid also produces m/z = 61 with $H_3O^+$, therefore m/z = 43 alone was used to measure propanol for this reagent ion. The $NO^+$ reagent ion was considered more reliable than $H_3O^+$, as the given products constitute the full list of products formed and the probability of products from other reactions displaying coincident masses is lower than for $H_3O^+$. However, as it adds only one more ion and may still provide useful information, the m/z = 43 ion was retained in the SIM scan method.

$$H_3O^+ + C_3H_7OH \rightarrow C_3H_7^+ (m/z = 43, 90\ \%) + 2H_2O \quad (4.21)$$

$$H_3O^+ + C_3H_7OH \rightarrow C_3H_7OH.H^+ (m/z = 61, 10\ \%) + H_2O \quad (4.22)$$

$$NO^+ + C_3H_7OH \rightarrow C_3H_7O^+ (m/z = 59) + HNO \quad (4.23)$$

$$C_3H_7O^+ + H_2O + M \rightarrow C_3H_7O^+.H_2O (m/z = 77) + M \quad (4.24)$$

$$C_3H_7O^+ + C_3H_7OH + M \rightarrow C_3H_7O^+.C_3H_7OH (m/z = 119) + M \quad (4.25)$$

### 4.3.5 Hexanal

Hexanal gives products with m/z = 101 and 119 from reaction with $H_3O^+$, (reactions 4.26 and 4.27) and m/z = 99 with $NO^+$ (reaction 4.28)(*3*). While both reagent ions can be used, the $H_3O^+$ products were not considered to add any useful information in this case, and were omitted in order to reduce the number of product masses scanned. Therefore, $NO^+$ alone was used for hexanal quantification.

$$H_3O^+ + C_5H_{11}CHO \rightarrow C_5H_{11}CHO.H^+ (m/z = 101) + H_2O \quad (4.26)$$

$$C_5H_{11}CHO.H^+ + H_2O + M \rightarrow C_5H_{11}CHO.H^+.H_2O (m/z = 119) + M \quad (4.27)$$

$$NO^+ + C_5H_{11}CHO \rightarrow C_5H_{11}CO^+ (m/z = 99) + HNO \quad (4.28)$$

### 4.3.6 Ethyl Acetate

As stated in section 3.4.2, ethyl acetate was highlighted by Morales et. al.(*12*) as a fermentation product. It forms m/z = 89 with the $H_3O^+$ reagent ion (reaction 4.29) and m/z = 43 and 118 with $NO^+$ (reactions 4.30 and 4.31)(*13*). Only m/z = 118 was included in the SIM scan method to measure ethyl acetate.

$$H_3O^+ + CH_3COOC_2H_5 \rightarrow CH_3COOC_2H_5.H^+ (m/z = 89) + H_2O \quad (4.29)$$

$$NO^+ + CH_3COOC_2H_5 + M \rightarrow CH_3COOC_2H_5.NO^+ (m/z = 118, 90\ \%) + M \quad (4.30)$$

$$NO^+ + CH_3COOC_2H_5 \rightarrow CH_3CO^+ (m/z = 43, 10\ \%) + C_2H_5NO_2 \quad (4.31)$$

There is no way to be sure if the products chosen here represent ethyl acetate exclusively. Methyl propanoate and butanoic acid both form products at m/z = 89 with $H_3O^+$ and m/z = 118 with $NO^+$. Butanoic acid also forms m/z = 71 with $NO^+$ and methyl propanoate forms m/z = 57, although products appearing at both of these masses are also formed by other compounds (butanal and propanal respectively), providing no routine check for the identity of the parent molecules in this instance. The deciding factors in the assignment of these product masses to ethyl acetate were the assertion of the presence of ethyl acetate in winey olive oil by Morales et. al.(*12*) through the use of dynamic head space gas chromatography and the observation of the appropriate product masses while measuring the head space above olive oil which displayed the winey defect. Butanoic acid is also a fermentation product, so differentiation between this and ethyl acetate is not crucial, as in most samples they are expected to be either both present or both absent. Methyl propanoate does not appear to have been identified in olive oil by any researchers to date. It is also not certain if it would be a fermentation product if present. For the purposes of this research the product peak observed at m/z = 118 from $NO^+$ was assigned to ethyl acetate. It should be recognised however that an olive oil should not be classified as having undergone fermentation based solely on

a high concentration of ethyl acetate – the presence of other fermentation products should also be confirmed.

### 4.3.7 (E)-2-Pentenal

(E)-2-pentenal was identified by Morales et. al.(*6*) as being associated with ripe and fruity sensory attributes. SIFT-MS analysis produces m/z = 85 with the $H_3O^+$ reagent ion (reaction 4.32) and m/z = 83 and 114 with the $NO^+$ reagent ion (reactions 4.33 and 4.34)(*14*). There are few other VOCs which give the same product masses as (E)-2-pentenal. The isomers of hexanol are the most prominent, although they only coincide with (E)-2-pentenal when the $H_3O^+$ reagent ion is used (also producing m/z = 85) and no $NO^+$ hexanol product masses were identified in olive oil during this research. m/z = 101 is the sole hexanol product with the $NO^+$ reagent ion. $NO^+$ alone was considered sufficient to measure (E)-2-pentenal. A Henry's law constant was not available for (E)-2-pentenal, so including it in the Henry's law study to verify the accuracy of the calculated concentrations was not possible.

$$H_3O^+ + C_2H_5CHCHCHO \rightarrow C_2H_5CHCHCHO.H^+ (m/z = 85) + H_2O \quad (4.32)$$

$$NO^+ + C_2H_5CHCHCHO \rightarrow C_2H_5CHCHCO^+ (m/z = 83, 95\ \%) + HNO \quad (4.33)$$

$$NO^+ + C_2H_5CHCHCHO + M \rightarrow C_2H_5CHCHCHO.NO^+ (m/z = 114, 5\ \%) + M \quad (4.34)$$

### 4.3.8 Pentanol isomers

As identified by Angerosa et. al.(*15*) and mentioned in section 3.4.3, 3-methyl butan-1-ol is a prominent product of olive fermentation which leads to the 'fusty' defect in the resultant oil. As mentioned previously, SIFT-MS is not able to adequately distinguish between different isomers of alcohols such as 1-pentanol(*1*) – this includes 3-methyl butan-1-ol. Therefore, in place of reporting 3-methyl butan-1-ol when there is little evidence that the detected analyte is in fact this isomer, all pentanol isomers were reported as a group. The m/z values assigned to pentanol were 71 with the $H_3O^+$ reagent ion (reaction 4.35) and m/z = 71 and 87 with $NO^+$ (reactions 4.36 and 4.37)(*1*). Measurements were taken including all product masses shown here, yet only the $H_3O^+$ results were included in subsequent calculations of concentration. There is variation in product ions formed between isomers when $NO^+$ is used, with 2-methyl-1-butanol and 2-methyl-2-butanol both producing exclusively m/z = 71. The product of butanal with $NO^+$ also appears at m/z = 71, so there is opportunity for interference at this mass. The m/z = 87 ion is a reliable confirmatory product with $NO^+$ for the majority of pentanol isomers, and it was used for confirmation of the $H_3O^+$ derived concentrations. The five available rate constants for reaction of pentanol isomers with $H_3O^+$ are all $2.8 \times 10^{-9}$ $cm^3$ $molecule^{-1}$ $s^{-1}$(*1*). This allows the total concentration of these pentanol isomers to be calculated.

$$H_3O^+ + C_5H_{11}OH \rightarrow C_5H_{11}^+ (m/z = 71) + 2H_2O \qquad (4.35)$$

$$NO^+ + C_5H_{11}OH \rightarrow C_5H_{11}O^+ (m/z = 87, 85\ \%) + HNO \quad (4.36)$$

$$NO^+ + C_5H_{11}OH \rightarrow C_5H_{11}^+ (m/z = 71, 5\ \%) + HNO_2 \quad (4.37)$$

### 4.3.9 Acetic Acid

Acetic acid produces peaks at m/z = 61, 79 and 97 with $H_3O^+$ (reactions 4.38, 4.39 and 4.40), and m/z = 90 with $NO^+$ (reaction 4.41)(*13*). The $H_3O^+$ peaks can coincide with those of the ethanol-methanol cluster, so $NO^+$ was considered to be the more reliable reagent ion. Besides the proton transfer product and associated water clusters, acetic acid can also cluster to itself at higher concentrations to produce peaks at m/z = 121 and 139 when $H_3O^+$ is used (reactions 4.42 and 4.43). These are characteristic of acetic acid at higher concentrations, and the m/z = 121 peak is very useful for confirmation in a mass scan. The m/z = 139 peak, however, coincides with the ethanol proton-bound triple cluster which is present at high concentrations of ethanol. This is a significant concern, as when there is a high concentration of acetic acid in an olive oil it is usually due to fermentation. As ethanol is the primary product of fermentation, it is also expected to be present at high concentration in these situations. Hence, after fermentation the m/z = 139 cluster will always be observed. As in the case of the methanol dimer cluster investigated in figures 4.3 and 4.4, the acetic acid contribution to the m/z = 139 peak in most samples is considered to be insignificant compared with that from ethanol. It should also be noted that when ethanol is

present in the flow tube at a high enough concentration to produce the m/z = 139 triple cluster, the $H_3O^+$ reagent ion is often depleted to such a degree that the calculated concentrations for analytes are inaccurate. Under these conditions the situation can be simply remedied by reducing the sample flow.

$$H_3O^+ + CH_3COOH \rightarrow CH_3COOH.H^+ (m/z = 61) + H_2O \quad (4.38)$$

$$CH_3COOH.H^+ + H_2O + M \rightarrow CH_3COOH.H^+.H_2O (m/z = 79) + M \quad (4.39)$$

$$CH_3COOH.H^+.H_2O + H_2O + M \rightarrow CH_3COOH.H^+.2H_2O (m/z = 97) + M \quad (4.40)$$

$$NO^+ + CH_3COOH + M \rightarrow CH_3COOH.NO^+ (m/z = 90) + M \quad (4.41)$$

$$CH_3COOH.H^+ + CH_3COOH + M \rightarrow (CH_3COOH)_2.H^+ (m/z = 121) + M \quad (4.42)$$

$$(CH_3COOH)_2.H^+ + H_2O + M \rightarrow (CH_3COOH)_2.H^+.H_2O (m/z = 139) + M \quad (4.43)$$

### 4.3.10 Propanoic Acid

Propanoic acid forms products at m/z = 57 and 75 with $H_3O^+$ (reactions 4.44 and 4.45) while it forms m/z = 57 and 104 with $NO^+$ (reactions 4.46 and 4.47)(*13*). The m/z = 57 and 75 products with $H_3O^+$ coincide with $^{18}O$ isotope products of the

$H_3O^+.2H_2O$ and $H_3O^+.3H_2O$ clusters. The $^{18}O$ isotope only makes up approximately 0.2 % of naturally occurring oxygen, however the water clusters concerned contain multiple oxygen nuclei (increasing the likelihood that one will be $^{18}O$), and the $^{16}O$ products at m/z = 55 and 73 are very abundant in the flow tube at atmospheric humidity. The proportion of these clusters contributing to the m/z = 57 and 75 signals may be calculated and removed, but the resulting concentrations will be less precise as a result. The better alternative is to use solely the $NO^+$ reagent ion.

$$H_3O^+ + C_2H_5COOH \rightarrow C_2H_5COOH.H^+ (m/z = 75, 90\ \%) + H_2O \quad (4.44)$$

$$H_3O^+ + C_2H_5COOH \rightarrow C_2H_5CO^+ (m/z = 57, 10\ \%) + 2H_2O \quad (4.45)$$

$$NO^+ + C_2H_5COOH + M \rightarrow C_2H_5COOH.NO^+ (m/z = 104, 70\ \%) + M \quad (4.46)$$

$$NO^+ + C_2H_5COOH \rightarrow C_2H_5CO^+ (m/z = 57, 30\ \%) + HNO_2 \quad (4.47)$$

The m/z = 57 product with $NO^+$ coincides with that of propanal, however this product represents only 30 % of propanoic acid products, while the propanal product is the only one formed. Propanal also reacts a full two-thirds faster with $NO^+$ than does propanoic acid (with a rate constant of $2.5 \times 10^{-9}$ versus $1.5 \times 10^{-9}$ $cm^3$ $molecule^{-1}$ $s^{-1}$). Therefore, propanoic acid contributes only one fifth as much to any signal observed at m/z = 57 as does propanal if both analytes are present at the same concentration. The peak at m/z = 104 gives a reliable measure

of propanoic acid which is independent of any products formed by propanal, as 70 % of all reactions between propanoic acid and $NO^+$ produce the m/z = 104 product. Only a high concentration of propanoic acid will have a significant effect on the calculated propanal concentration, and the m/z = 104 product will show when this is the case.

### 4.3.11 Remaining Unassigned Mass

After identifying the aforementioned compounds, there was left a single product mass which could not be explained. This was the m/z = 128 peak with the $NO^+$ reagent ion. The m/z = 128 peak may be due to either a cluster of identified VOCs with each other or an unexpected fragment. Nevertheless, this mass was measured so as to allow for retrospective calculation of a VOC concentration if it were assigned at a later date.

## 4.4 Concluding Remarks

The product-to-analyte assignments shown in this chapter allowed the construction of standardised SIM scan methods for analysis of olive oil VOCs. Using SIFT-MS, all thirteen analytes listed in this chapter may be quantified in an olive oil head space sample to acceptable precision in one minute. The studies performed using these SIM methods are described in chapters 5 and 6.

## 4.5 References


(1) Španel, P.; Smith, D. SIFT Studies of the Reactions of $H_3O^+$, $NO^+$ and $O_2^+$ with a Series of Alcohols. *Int. J. Mass Spectrom. Ion Proc.* **1997**, *167/168*, 375-388.

(2) Compilation of Henry's Law Constants for Inorganic and Organic Species of Potential Importance in Environmental Chemistry. http://www.mpch-mainz.mpg.de/~sander/res/henry.html (accessed Apr 20, 2007)

(3) Španel, P.; Ji, Y.; Smith, D. SIFT Studies of the Reactions of $H_3O^+$, $NO^+$ and $O_2^+$ with a Series of Aldehydes and Ketones. *Int. J. Mass Spectrom. Ion Proc.* **1997**, *165/166*, 25-37.

(4) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 639-660.

(5) Sánchez, J.; Harwood, J. L. Biosynthesis of Triacylglycerols and Volatiles in Olives. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 564-573.

(6) Morales, M. T.; Alonso, M. V.; Ríos, J. J.; Aparicio, R. Virgin Olive Oil Aroma: Relationship between Volatile Compounds and Sensory Attributes by Chemometrics. *J. Agric. Food Chem.* **1995**, *43*, 2925-2931.

(7) Kiritsakis, A. Flavor Components of Olive Oil - A Review. *J. Amer. Oil Chem. Soc.* **1998**, *75*, 673-681.

(8) Španel, P.; Smith, D. Selected Ion Flow Tube Studies of the Reactions of $H_3O^+$, $NO^+$ and $O_2^+$ with Several Aromatic and Aliphatic Hydrocarbons. *Int. J. Mass Spectrom.* **1998**, *181*, 1-10.

(9) Schoon, N.; Amelynck, C.; Vereecken, L.; Arijs, E. A Selected Ion Flow Tube Study of the Reactions of $H_3O^+$, $NO^+$ and $O_2^+$ with a Series of Monoterpenes. *Int. J. Mass Spectrom.* **2003**, *229*, 231-240.

(10) Reiners, J.; Grosch, W. Odorants of Virgin Olive Oils with Different Flavor Profiles. *J. Agric. Food Chem.* **1998**, *46*, 2754-2763.

(11) Grosch, W. Evaluation of the Key Odorants of Foods by Dilution Experiments, Aroma Models and Omission. *Chem. Senses* **2001**, *26*, 533-545.

(12) Morales, M. T.; Luna, G.; Aparicio, R. Sensory and Chemical Evaluation of Winey-Vinegary Defect in Virgin Olive Oils. *Eur. Food Res. Technol.* **2000**, *211*, 222-228.

(13) Španel, P.; Smith, D. SIFT studies of the reactions of $H_3O^+$, $NO^+$ and $O_2^+$ with a series of volatile carboxylic acids and esters. *Int. J. Mass Spectrom. Ion Proc.* **1998**, *172*, 137-147.

(14) Španel, P.; Van Doren, J.; Smith, D. A Selected Ion Flow Tube Study of the Reaction of $H_3O^+$, $NO^+$ and $O_2^+$ with Saturated and Usaturated Aldehydes and Subsequent Hydration of the Product Ions. *Int. J. Mass Spectrom.* **2002**, *213*, 163-176.

(15) Angerosa, F.; Di Giacinto, L.; Solinas, M. Influenza dello Stoccaggio in Massa delle Olive sull'Aroma degli Oli di Risulta: Valutazione del Difetto di "Riscaldo" Mediante Analisi HPLC e GLC dei Componenti Volatili. *Riv. Merceol.* **1990**, *29*, 275-294.

# Chapter 5

## Olive Oil Defects

### 5.1 Introduction

Samples of olive oils displaying different sensory defects were donated by both the organisers of the Olives New Zealand tasting panel at Hort Research in Auckland, New Zealand, and the International Olive Oil Council in Madrid, Spain. These were invaluable, as they allowed the identification of the appropriate VOC products which could be detected using SIFT-MS. To investigate each defect in-depth, however, requires the recreation of the conditions under which the defect arises in fresh oil. The experiments and attempts to produce these defects are the subjects of the present chapter. The sensory defects known as 'rancid', 'winey', 'musty', 'fusty' and 'muddy' are discussed.

### 5.2 Rancid Olive Oil

A sample of olive oil displaying the rancid sensory defect was donated by the organisers of the Olives New Zealand tasting panel at Hort Research in Auckland. 10 mL oil was placed in a 125 mL glass vial. The vial was purged with a constant stream of nitrogen gas during SIFT-MS analysis. A mass scan using the $NO^+$ reagent ion was initially performed, from which the mass spectrum shown in figure 5.1 was obtained.

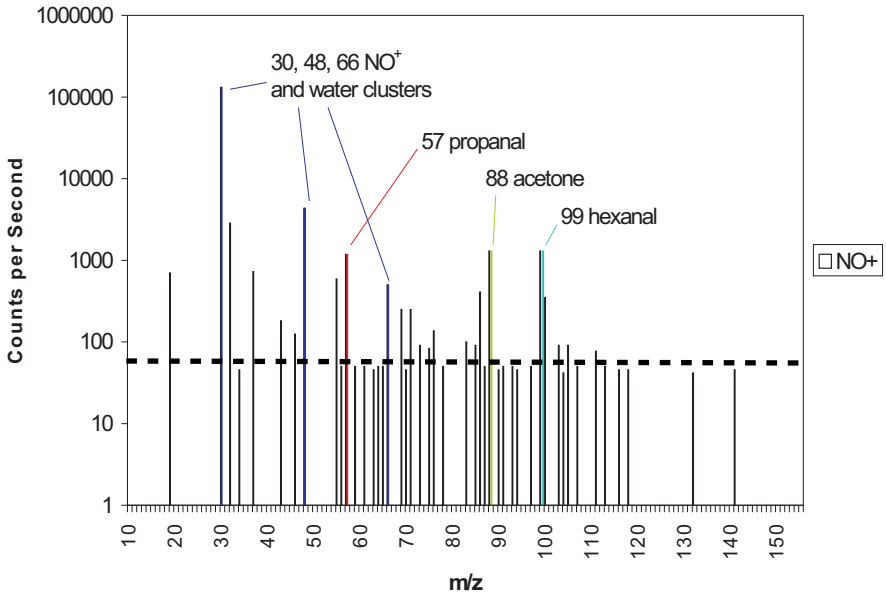



**Figure 5.1.** A mass spectrum of a rancid standard olive oil using the $NO^+$ reagent ion. This scan was obtained from olive oil head space in just over 17 seconds. Note the major mass peaks at m/z = 57, 88 and 99 corresponding to products of propanal, acetone and hexanal respectively. Only the major oxidation products were observed as the short dwell time at each mass caused a high baseline count level which obscured lower concentration VOCs. Note the logarithmic scale on the vertical axis.

The most noticeable features in the spectrum of rancid olive oil when compared with defect-free olive oil are the large peaks which appear at m/z = 57, 88 and 99 when using the $NO^+$ reagent ion. These product peaks correspond to propanal, acetone and hexanal respectively. Propanal, according to Belitz and Grosch(*1*), is the major autoxidation product of linolenic acid, while hexanal is the second-most major product of linoleic acid autoxidation, after pentane. Acetone is

not mentioned by Belitz and Grosch, nor indeed by many researchers. Propanal is also not a widely reported oxidation product of olive oil, however it is recognised as a major oxidation product of oils with a high concentration of linolenic acid, such as fish oils(*2;3*). Both acetone and propanal have been reported by Angerosa(*4*), Procida et. al.(*5*) and Solinas et. al.(*6*) as being present in olive oil, however only Solinas et. al. were investigating oil oxidation. Acetone and propanal were both observed to increase in concentration with an increasing degree of rancidity. Curiously, in a follow-up article by Solinas et. al.(*7*) where the products of oxidation were quantified, neither acetone nor propanal were mentioned.

The production of pentane, as reported by Belitz and Grosch(*1*), was not investigated, as the analysis of pentane by SIFT-MS is complicated and unreliable at present. Pentane does not react with $H_3O^+$ or $NO^+$, and produces very similar products to other hydrocarbons with $O_2^+$(*8*). The peaks of the $O_2^+$ spectrum are very difficult to assign when more than one hydrocarbon is present, as all hydrocarbons larger than ethane are fragmented and many give peaks at m/z = 42, 43, 57 and 72, the masses observed from reaction of $O_2^+$ with pentane. Many hydrocarbons can be distinguished only by the highest mass peaks they produce (the electron transfer product with the same mass as the parent molecule) and the proportions of each mass fragment. Proportions of these fragments for each individual compound are difficult to discern when there are several compounds contributing to each peak. There are enough products from other compounds whose masses coincide with those of pentane's products to prohibit reliable quantitation, and in most cases even reliable identification.

The corresponding mass scan of the rancid standard oil using the $H_3O^+$ reagent ion is shown in figure 5.2. Many more peaks are present above the background noise level than when the $NO^+$ reagent ion is used, due to the much more prevalent formation of water clusters for $H_3O^+$. Ethanol and methanol are both present at approximately typical concentrations for an olive oil. Acetone and propanal (which form products of identical mass from reaction with $H_3O^+$) have large peaks at m/z = 59, 77 and 95. This is in good agreement with figure 5.1. Hexanal forms products at m/z = 83 and 101 in approximately equal proportions with $H_3O^+$. The m/z = 83 peak from hexanal coincides with an ethanol water cluster ($C_2H_5OH.H^+.2H_2O$), so will not provide accurate results. The m/z = 101 peak is at the level of the background noise and would not be noticed unless sought out specifically. The case of hexanal here illustrates well the benefits of employing multiple precursor ions. One compound conspicuous by its absence from both figures 5.1 and 5.2 is (E)-2-hexenal. It is always one of the most prominent VOCs observed in high quality olive oil. (E)-2-hexenal is not expected to be observed at high concentrations in rancid olive oil, and the present sample suggests that this important compound is not present at all.

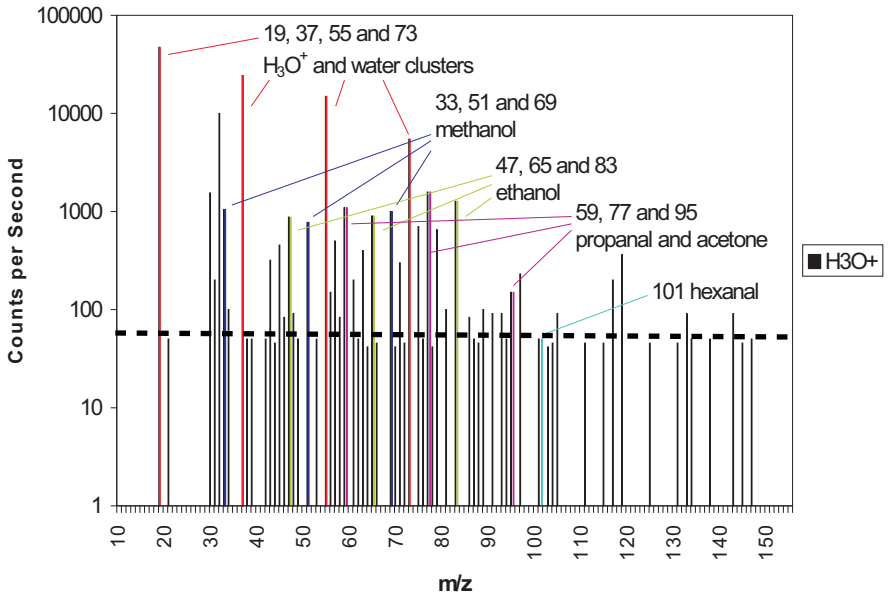



**Figure 5.2.** A mass spectrum of a rancid standard olive oil using the $H_3O^+$ reagent ion. The product ions observed here aid in identification of those in figure 5.1.

The preceding work permitted the identification of the rancid defect in olive oil from the observation of acetone, propanal and hexanal. The present research requires that quantification also be achieved. Several different samples of olive oil displaying different degrees of rancidity were needed so that the correlations between the degree of rancidity and the concentrations of VOCs could be discovered. The most appropriate experiment scheme consists of driving fresh olive oil to rancidity and analysing the oil at selected intervals during the process. Pilot studies were conducted until the most appropriate experimental method could be defined.

For the first pilot study, 50 mL of a New Zealand olive oil were placed in each of seven 125 mL glass vials and heated in an oven at 70 ± 4 °C for two weeks.

One bottle was removed every two days over the course of the two weeks for direct head space analysis by SIFT-MS. A large increase was observed in the concentrations of acetone and propanal over this period. Though the oil was very likely rancid by the end of the experiment, the degree of rancidity is not known, nor at which point it became rancid. This experiment was useful in that it exposed the important products of olive oil oxidation and showed the timescale over which these products are formed. However the conditions under which it was conducted were too variable (the oven thermostat was not satisfactory, varying by ± 5 °C), and no independent measure of oxidation was used to compare with the SIFT-MS results. At least one established measure of oil oxidation was necessary for comparison if any useful information was to be obtained from SIFT-MS analysis.

It was decided to use two standard methods: sensory testing by a trained panel and a peroxide test by a commercial laboratory. A heated water bath maintained at 60 °C was used in place of the oven. To determine the rate at which the oils would oxidise under these new conditions, 50 mL of two New Zealand olive oils were each sealed in a 125 mL glass vial and heated at 60 ± 1 °C in a water bath for approximately one week. All known head space VOCs were monitored over that time. As this was a preliminary study, several compromises were made. There was not sufficient oil available to fill several bottles, therefore each analysis was performed on only a single vial. The use of a single vial for each sample introduced the largest possibility for discrepancy between this pilot study and the subsequent experiment. Each vial was sealed with an airtight septum and needed to be purged with air after analysis to return it to atmospheric pressure. If the pressure

were not restored, the pressure inside the vial would be lowered with each subsequent analysis. This would lower the sample flow rate into the flow tube – which is dependent on the difference in pressure between the flow tube and the sample. A lower sample pressure results in a lower sample flow and therefore a lower analyte concentration, causing the value reported to be lower than the true value.

The re-pressurisation issue would not exist in the final experiment, as each vial would be removed for analysis and not re-introduced to the water bath. Its importance lay in its effect on the rate of oil oxidation. The foremost objective of the pilot studies was to obtain an estimation of the time required under the chosen conditions for the oil to produce significant concentrations of oxidation products. Both pilot studies led to the belief that two weeks would be appropriate for fresh olive oil to become rancid.

Two litres of a New Zealand olive oil from a local producer were obtained for use in the experiment proper. A 50 mL aliquot was placed in each of 29 125 mL glass vials, sealed with a rubber septum cap, placed in a water bath at 60 ± 1 °C and kept in the dark. The remaining oil was subjected to SIFT-MS VOC analysis, a peroxide test performed by AgriQuality in Auckland, and a taste test performed by the Olives New Zealand sensory panel, also in Auckland. One vial was removed from the water bath each day, the head space purged with nitrogen and the vial stored in a cool, dark cupboard until SIFT-MS analysis. Every day the VOCs were measured using SIFT-MS and every second day the peroxide value was also determined. Every four days, five vials were removed in addition to that for

SIFT-MS analysis and sent to the Olives New Zealand sensory panel for determination of the strength of the rancid sensory defect. The experiment was planned for two weeks, yet on analysis of the samples by SIFT-MS after one week, the oil was not considered to be oxidising rapidly enough to display a noticeable rancid defect by the end of the two weeks. Therefore, the sample removal frequency was halved to every two days, and the timing of all analyses altered accordingly.

This experiment was repeated with two other New Zealand olive oils. However, both of these were oxidised for only two weeks, in accordance with the original experiment plan.

SIFT-MS analysis involved a SIM scan of 14 VOCs: methanol, ethanol, propanol, pentanol, propanal, acetone, hexanal, (Z)-3-hexenal, (E)-2-hexenal, ethyl acetate, acetic acid, propanoic acid, (E)-2-pentenal and nonanal. The correct quantification of several important VOCs was confirmed by the Henry's Law studies reported in section 4.3. On no occasion was nonanal found above the background level and therefore results for only 13 VOCs were used for Partial Least-Squares (PLS) regression.

Results for all three oils from the peroxide value and sensory rancidity determinations are shown in figures 5.3 and 5.4. There was an increase in both parameters over time for the three oils. All three oils displayed a linear relationship between time of oxidation and peroxide value (i.e. a linear function gave a higher $R^2$ value than other common function types applied), albeit with different coefficients (figure 5.3). The progression of rancidity over time for the three oils

was not able to be monitored as closely as that for peroxide value, with only twelve measurements possible. Oil A showed a slowing rate of rancidity over the course of the experiment, while oils B and C became rancid in a more linear fashion (figure 5.4).

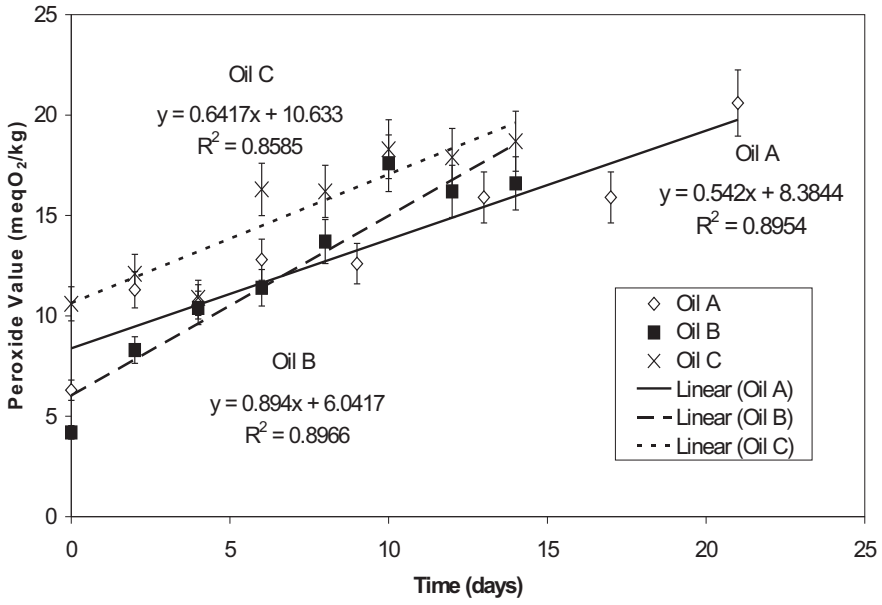



**Figure 5.3.** The increase of peroxide value with time for the three oils included in the oxidation study. Values have an error of ± 8 %.

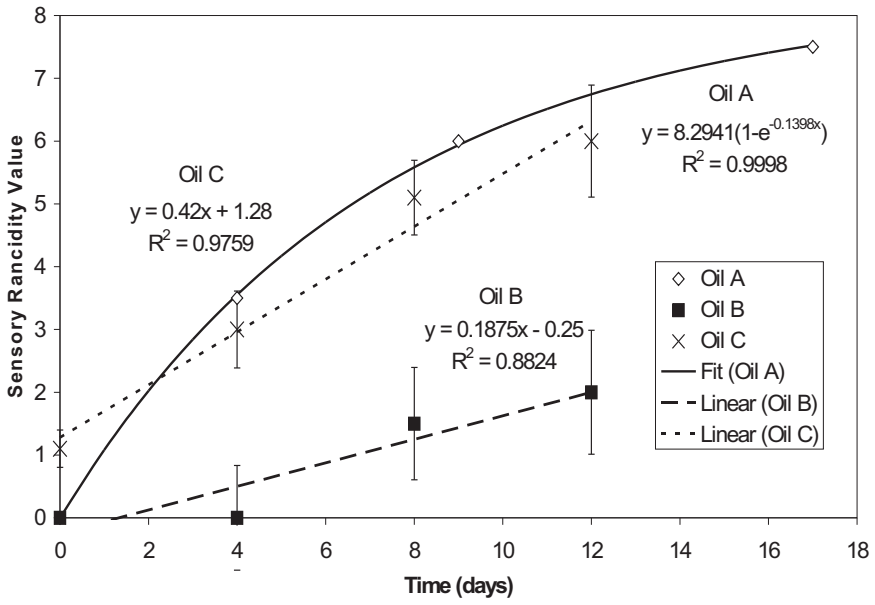
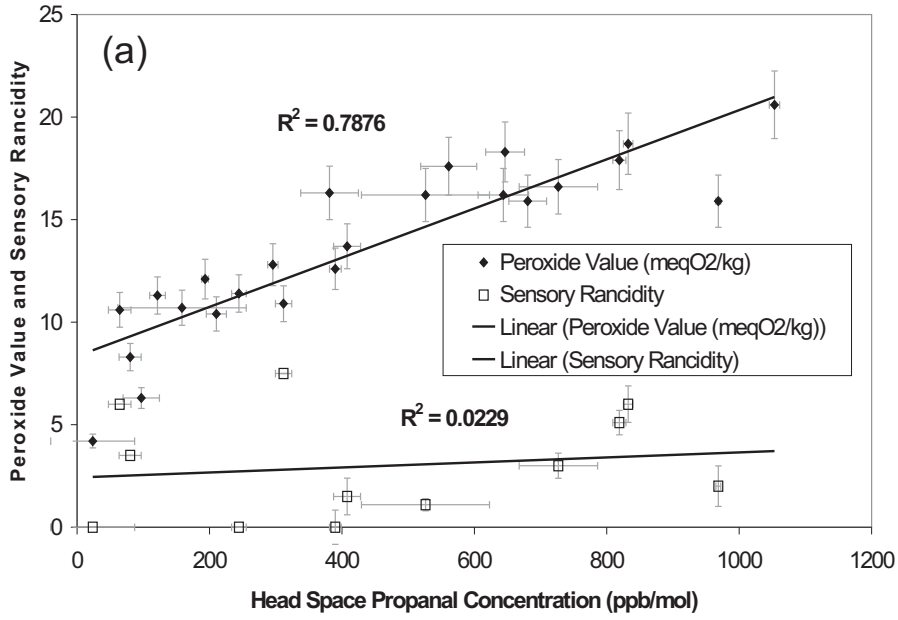
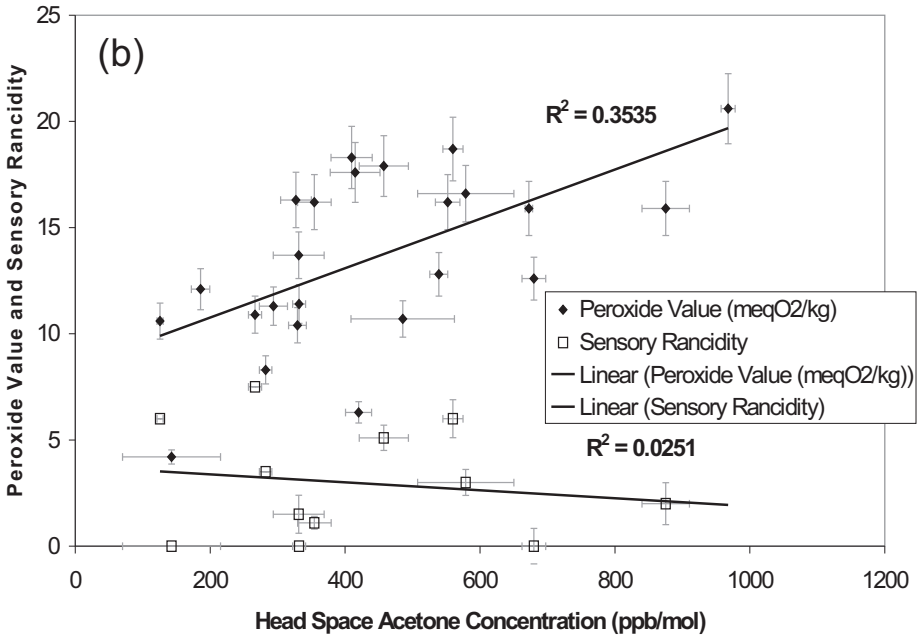



**Figure 5.4.** The increase of sensory rancidity value with time for the three oils included in the oxidation study. Oil A shows a non-linear relationship which is coincidentally best modelled by an inverted exponential decay function (see section 8.2.2). Sensory rancidity values are medians, errors are expressed as robust standard deviations(*9*). Errors for oil A were not available.

Only three VOCs were observed to increase uniformly in all three oils during this experiment: propanal, acetone and acetic acid figure 5.5. Hexanal was expected to increase along with these compounds, yet no change was seen in its concentration in any oil head space. Nonanal was also not observed, although this is less surprising as it was not detected in the head space above the rancid standard obtained from the IOOC and is not considered volatile enough to be present at appreciable concentrations in all but the most rancid olive oils – even then it may need elevated temperature to be measured. As suggested by the results for other

olive oil defects below, acetone and acetic acid are produced by several different processes; the production of propanal seems to be specific to rancidity.



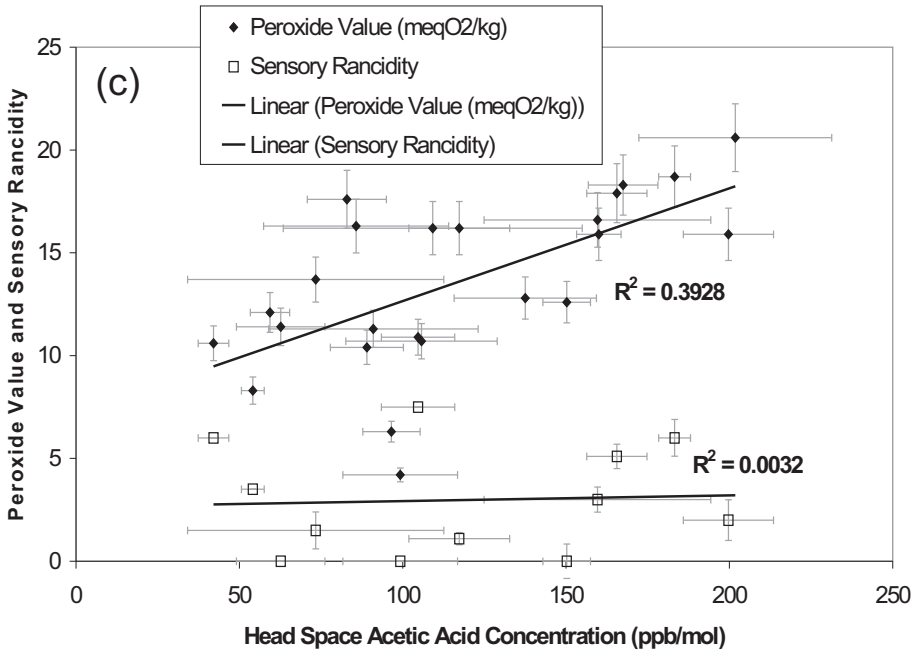



**Figure 5.5.** Correlations of peroxide value and sensory rancidity with head space concentrations of **(a)** propanal, **(b)** acetone and **(c)** acetic acid.

When compared with the rise in peroxide value, propanal showed a very good correlation for all three oils, with an overall correlation coefficient ($R^2$) of 0.79. An improved fit was obtained with a logarithmic function ($R^2$ = 0.85), although there seems to be no reason why a logarithmic function should achieve a better fit. There was a poor correlation between the head space propanal concentration and the sensory rancidity observed for all three oils together ($R^2$ = 0.02). However, propanal increased with oxidation time and rancidity behaved likewise, causing the two parameters to display similar behaviour for each oil individually. For oils B and C the individual linear correlations, while different from each other, were significant ($R^2$ = 0.89 and 0.95 respectively). Oil A, on the other hand, was closer to displaying a logarithmic correlation ($R^2$ = 0.91) with

propanal concentration than a linear one ($R^2$ = 0.71, figure 5.6). With oils B and C displaying a linear relationship, there seems to be no reason why oil A should differ. This is assumed to be due to a discrepancy in the tasting data, as the taste test is the least repeatable technique employed in this study.

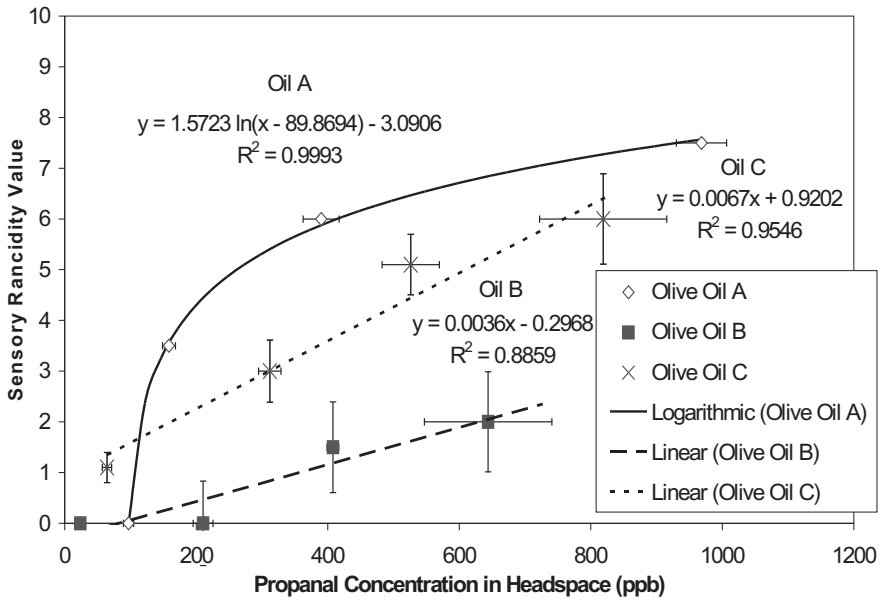



**Figure 5.6.** Individual correlations between sensory rancidity values of the three oils and the head space propanal concentration. The results for oil A are distinctly non-linear (approximated here by a logarithmic function), while the other two oils show good linear relationships.

Due to the fewer available results and lack of agreement between oils for the sensory data, it was not considered worthwhile to use PLS regression for the prediction of sensory rancidity. The peroxide value data were considered to be more reliable and much better suited to multivariate regression techniques. Subsequently, these results are described here.

There are two measures used to assess the results of PLS regression: cross-validation and cross-verification. Each is used for a different purpose. Cross-validation is used to determine how complicated the data are and decide how many latent variables are needed to provide the best prediction. Cross-verification is used to evaluate the prediction on a new set of data. It determines how closely the calibration samples approximate the population and is used to discover how well future samples will be predicted. Section A.3 has more information on this.

Leave-one-out cross-validation was performed to obtain an estimate of the error of the predicted peroxide value. This is a common approach when only a limited number of oils are available. Although, as noted in section A.3, it may give an over-optimistic appraisal if the sample is not representative of the population to which the model is to be applied.

An independent test set was not used for cross-verification, as the data set was small and all data were necessary for calibration. Foregoing independent cross-verification is suggested by Martens and Dardenne(*10*) for small data sets. The relevant output data from PLS regression are shown in table 5.2. These are the RMSECV value – a measure of the prediction error when different numbers of latent variables are included, the $R^2$ value – a measure of the correlation between the measured and predicted peroxide values, the **P** matrix – a measure of the importance of each VOC to each latent variable and the $\hat{\mathbf{y}}_{\mathbf{un}}$ vector – the peroxide values predicted for each oil sample from leave-one-out cross-validation. Each of the cross-validation measures (RMSECV and $R^2$) suggests a different number of latent variables be included. The RMSECV and $R^2$ values are at their optimum

values (RMSECV should be minimised, $R^2$ should be maximised) for a model with three and four latent variables respectively. Three latent variables were included in the model, as the increase in $R^2$ value from three to four latent variables was less than 0.002 and hence did not provide a significant improvement in predictive performance.

**Table 5.2.** Output from PLS regression. The rows of RMSECV and $R^2$ correspond to the latent variables (LV), the rows of the **P** matrix correspond to the VOCs (listed), while the columns correspond to the latent variables (also listed) and the rows of the $\hat{\mathbf{y}}_{\mathbf{un}}$ vector correspond to the oils analysed. RMSECV and $R^2$ values suggest the use of three latent variables, as the fourth gives only a slight improvement. Propanal, acetic acid and acetone display the highest loadings on the first latent variable (LV), therefore are the most important VOCs for prediction of peroxide values. See the text and section A.3 for more information.

| RMSECV | $R^2$ | **P** | | | | $\hat{\mathbf{y}}_{\mathbf{un}}$ |
|---|---|---|---|---|---|---|
| | | VOC | 1st LV | 2nd LV | 3rd LV | |
| 2.8 | 0.522 | Methanol | -0.108 | -0.408 | 0.161 | 8.26 |
| 2.06 | 0.745 | Ethanol | -0.202 | -0.317 | 0.453 | 10.6 |
| 1.61 | 0.841 | Propanol | 0.0146 | -0.416 | 0.308 | 11.1 |
| 1.6 | 0.842 | Propanal | 0.544 | -0.011 | 0.331 | 10.3 |
| 1.92 | 0.793 | Acetone | 0.47 | -0.206 | 0.189 | 13.3 |
| 1.89 | 0.794 | Hexanal | 0.107 | -0.362 | 0.0026 | 14.5 |
| 1.86 | 0.794 | (Z)-3-hexenal | 0.0073 | -0.362 | 0.291 | 17.9 |
| 2.08 | 0.754 | (E)-2-hexenal | 0.201 | -0.267 | -0.31 | 18.6 |
| 2.4 | 0.7 | Ethyl acetate | 0.224 | 0.154 | -0.345 | 4.93 |
| 2.4 | 0.705 | (E)-2-pentenal | 0.313 | -0.351 | 0.0978 | 11.5 |
| 2.36 | 0.711 | Acetic acid | 0.489 | -0.201 | 0.218 | 9.65 |
| 2.31 | 0.721 | Propanoic acid | 0.351 | -0.228 | -0.514 | 12.9 |
| 2.37 | 0.709 | Pentanol isomers | 0.257 | -0.37 | 0.0583 | 13.9 |
| | | | | | | 15.2 |
| | | | | | | 15 |
| | | | | | | 19.1 |
| | | | | | | 9.41 |
| | | | | | | 11.4 |
| | | | | | | 13.1 |
| | | | | | | 14.1 |
| | | | | | | 15.4 |
| | | | | | | 17.2 |
| | | | | | | 18.2 |
| | | | | | | 19.8 |

Plotting the $\hat{\mathbf{y}}_{\mathbf{un}}$ vector against the original **y** vector (see section A.3) allows evaluation of the fit from leave-one-out cross-validation. The $R^2$ value of the

comparison is 0.84, as listed in the $R^2$ vector which was used to suggest how many latent variables to include. Figure 5.7 is a visual interpretation of the process being carried out in the production of the vector of $R^2$ values. Plotting these vectors against each other provides a good check that the model is performing as expected. It must be remembered that a high $R^2$ value does not necessarily mean that the two variables are equal, it simply means they are correlated.

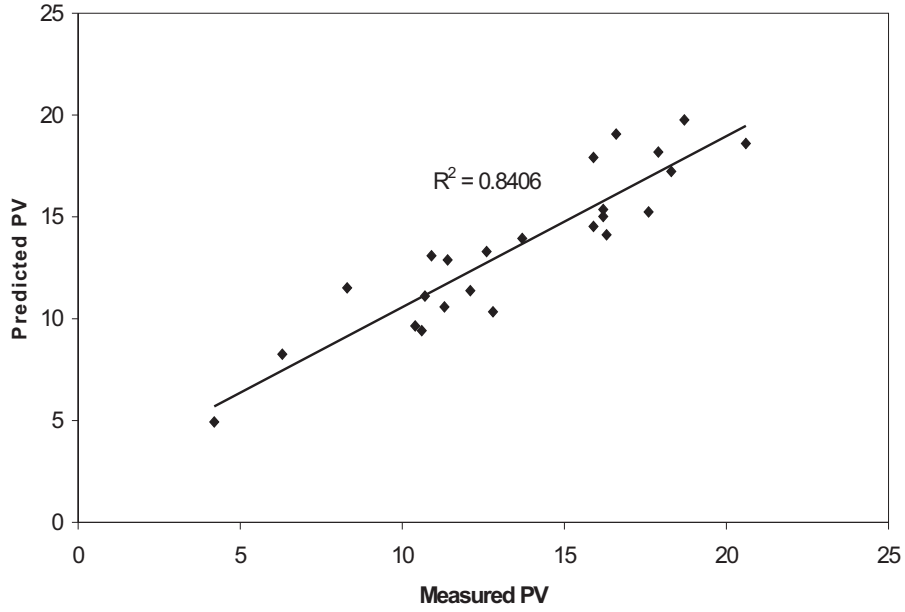



**Figure 5.7.** Measured versus predicted peroxide values for the 24 oil samples from "leave-one-out" cross-validation. Three latent variables were included.

From the original three oils exposed to oxidation, eight samples were taken of each oil making a total of 24 samples exposed for varying amounts of time to oxidation. The measured peroxide values for each oil are plotted against the predicted peroxide value in figure 5.7 using the model of table 5.2. To determine which VOCs have been identified as important by the PLS algorithm, the latent

variable loadings matrix is consulted. None of the loadings are very high – the highest in absolute magnitude is 0.54 and this is one of only two loadings above 0.5 in either direction on any latent variable. The loadings on the first latent variable are the most important, as they describe the single most important correlation between the VOC concentrations and peroxide values. The other latent variables are useful for prediction, but the physical relevance of each is unknown. As described in section A.3, there are no clear-cut rules about selecting which loading values are significant. For these results, $\pm 0.40$ was arbitrarily chosen. In reality this is possibly too low, however incorrectly considering certain loadings to be significant does not alter the prediction capability. Interpreting the loadings is an optional practice which may permit an understanding of the relationships between the independent and dependent variables, it is by no means necessary in PLS. Propanal, acetone and acetic acid display the highest correlations with the first latent variable. The loading of propanal on this first and most important latent variable is the highest of any VOC on any latent variable, which shows that propanal is an important oxidation product. The other latent variables do not demonstrate any clear pattern, so do not provide any additional insight.

Prediction errors, calculated to 95 % confidence as outlined in section A.3, are presented in figure 5.8 for the three different oils exposed to the oxidation procedure described earlier. These values provide a good indication of the precision which may be expected from the prediction of peroxide values from measuring VOCs.

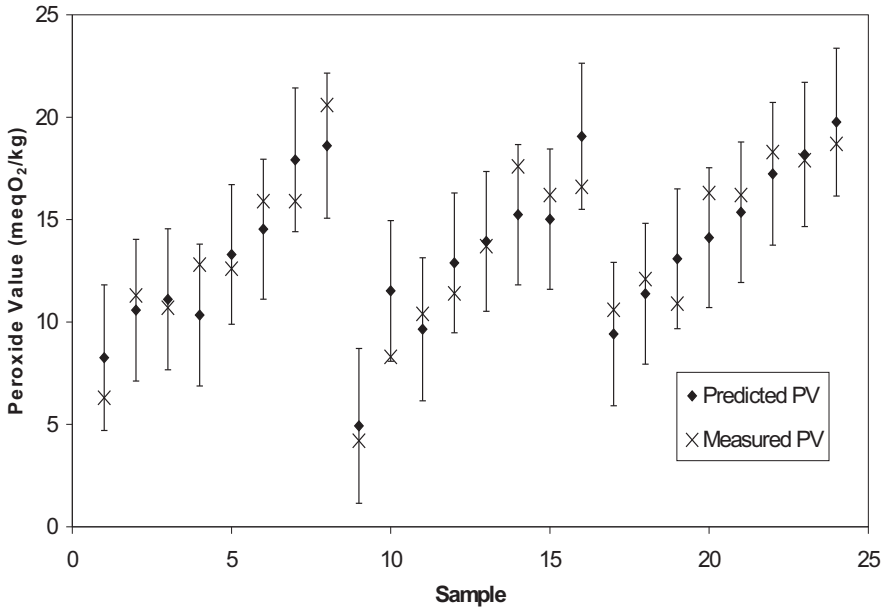



**Figure 5.8.** A graphical representation of the prediction of peroxide values for eight samples of three different oils oxidised for different lengths of time obtained from leave-one-out cross-validation. The predicted values are shown along with the corresponding measured values and the prediction intervals at the 95 % confidence level.

To perform subsequent prediction of peroxide values, the measured concentrations of all VOCs must be scaled by subtracting the mean of the VOC concentrations from the original data and dividing by the standard deviation. Once scaled, this matrix is postmultiplied by the **b** vector (see section A.3). The resultant vector is multiplied by the standard deviation of the original peroxide values and added to the mean to give the predicted peroxide values. All necessary quantities mentioned here are given in table 5.4.

**Table 5.4.** The **b** vector derived from PLS regression, means and standard deviations of input variables. Values are given to four significant figures to prevent round-off error during prediction.

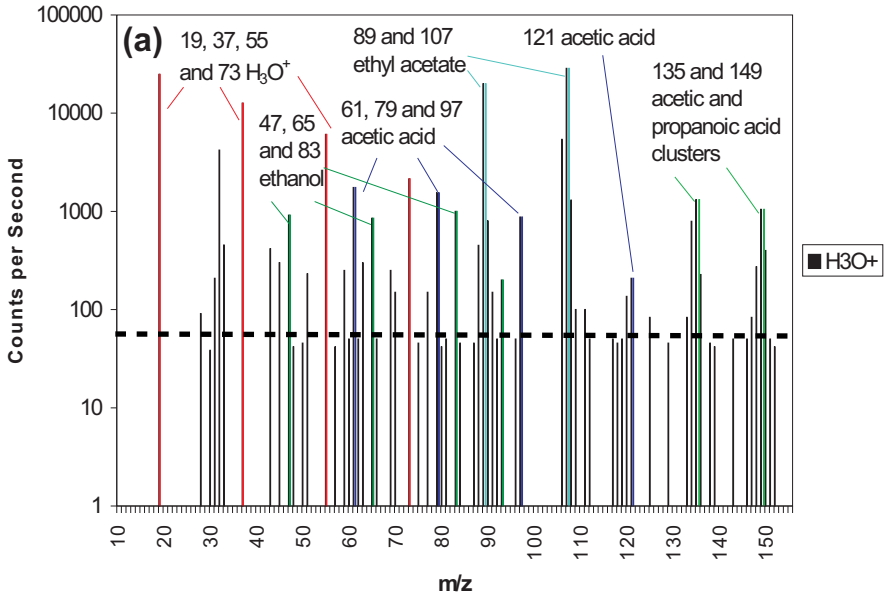
| **b** | Variable | Mean | Std Dev |
|---|---|---|---|
| -0.2338 | Methanol | 10930 | 3948 |
| -0.1192 | Ethanol | 4845 | 1505 |
| -0.0822 | Propanol | 79.71 | 27.06 |
| 0.4996 | Propanal | 434.9 | 304.8 |
| 0.2566 | Acetone | 440.9 | 211 |
| -0.0712 | Hexanal | 58.16 | 14.17 |
| -0.0823 | (Z)-3-hexenal | 230.4 | 58.64 |
| -0.0695 | (E)-2-hexenal | 274.4 | 163.3 |
| -0.0128 | Ethyl acetate | 32.59 | 13.12 |
| 0.0668 | (E)-2-pentenal | 57.74 | 19.86 |
| 0.2739 | Acetic acid | 116.6 | 47.05 |
| -0.1424 | Propanoic acid | 53.08 | 22.01 |
| -0.0025 | Pentanol isomers | 50.31 | 12.46 |
| | Peroxide value | 13.56 | 4.121 |

As this regression was conducted using data from three different New Zealand olive oils, it is expected to be applicable to any New Zealand olive oil for the determination of peroxide value. Thorough testing is necessary, however, as only limited data were available (24 measurements due to oil volume constraints, where at least 50 would have been desirable) and the oils in this study were oxidised under controlled conditions (i.e. 60 °C in the dark). Oils oxidised under different conditions may display different relationships between VOCs and peroxide value, as peroxides break down at different rates under different conditions(*11*).

## 5.3 Winey Olive Oil

A sample of winey olive oil donated by the Olives New Zealand sensory panel organisers displayed many different product mass peaks. Almost all high

intensity peaks were due to ethanol, acetic acid or ethyl acetate, all fermentation products which were present at high concentrations and hence underwent secondary reactions during SIFT-MS analysis. These secondary ion-molecule reactions are the result of the product ions from each of the three analytes ethanol, acetic acid and ethyl acetate reacting either with their parent analytes or with each other. For example, the major product of the reaction of $H_3O^+$ with ethanol occurs at m/z = 47 ($C_2H_5OH.H^+$). This product ion then reacts with $H_2O$ to form $C_2H_5OH.H^+.H_2O$ at m/z = 65 and so on. Mass scans obtained from winey olive oil are shown in figure 5.9.



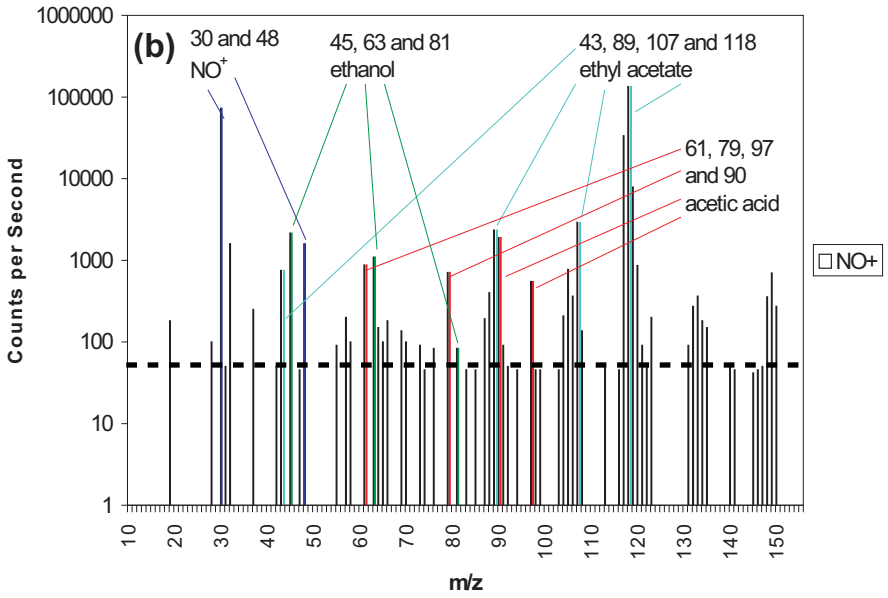



**Figure 5.9.** SIFT-MS mass scans of the head space above winey olive oil using **(a)** the $H_3O^+$ and **(b)** $NO^+$ reagent ions.

Clustering reactions of acetic acid and ethyl acetate with water proceed via identical mechanisms to those reactions which form methanol clusters as shown in reaction 4.3. Figure 5.9.(b) contains some abnormal mass peaks for an $NO^+$ spectrum. On occasion some metastable $NO^+$* can be present which generates $H_3O^+$. This $H_3O^+$ then reacts with the analytes present – its characteristic products with ethyl acetate and acetic acid may be observed here at m/z = 89 and 107 and m/z = 61, 79 and 97 respectively. The presence of small amounts of $H_3O^+$ can hamper the analysis of all but the highest concentration VOCs. Fortunately, the presence of low concentrations of other reagent ions in the flow tube is only important for samples containing VOCs at very high concentrations such as that shown in figure 5.9, as the contaminant reagent ions are always at very low

concentrations. In situations where several VOCs are present at very high concentrations and produce these unwanted products at detectable levels, identification of the product ions is straightforward, providing care is used in the interpretation of results. Quantification is not affected by the presence of product masses from other reagent ions as concentrations of VOCs high enough to give these products also significantly deplete (to less than 80 % of its original value) the reagent ion number density in the flow tube. Under these conditions, the approximation used to calculate VOC concentrations no longer holds and no calculated concentrations may be considered accurate. The problem of over-depletion of the reagent ion is easily solved by lowering the inlet flow of sample into the flow tube.

Investigation into determining the strength of the winey defect was also performed. However, due to the difficulty of re-creating the conditions necessary to produce the winey defect, this investigation was not able to be pursued to the same degree as the rancidity investigation. Only one sample of winey olive oil was available, so this was mixed with refined olive oil in varying proportions to produce oils which displayed different strengths of the winey defect. Several different methods are available to measure the degree of oxidation of olive oil, however a sensory test is the only established method for the evaluation of the other sensory defects, including the winey defect. Insufficient oil was available to supply the preferred sensory panel based at Hort Research in Auckland with the required oil volume (250 mL) for samples. This necessitated the use of a group of local tasters with limited training and experience, hence the results from this investigation are

not as reliable as those of the rancidity study. The method of diluting defective oil in non-defective oil is also not the ideal option, as the VOC profiles at different stages of defect development are much more likely to be accurate when samples are created in the same way as their real-world counterparts.

Winey olive oil obtained from the IOOC was added to refined olive oil in the following percent concentrations by volume: 50, 25, 12.5, 6.3, 3.1, 1.6 % and 0. The same 13 VOCs as in the rancidity study were measured by SIFT-MS. As the refined olive oil contributed only very low concentrations of VOCs, the concentrations of all identified species were found to decrease proportionally to the amount of refined oil added. These results are shown in figure 5.10.

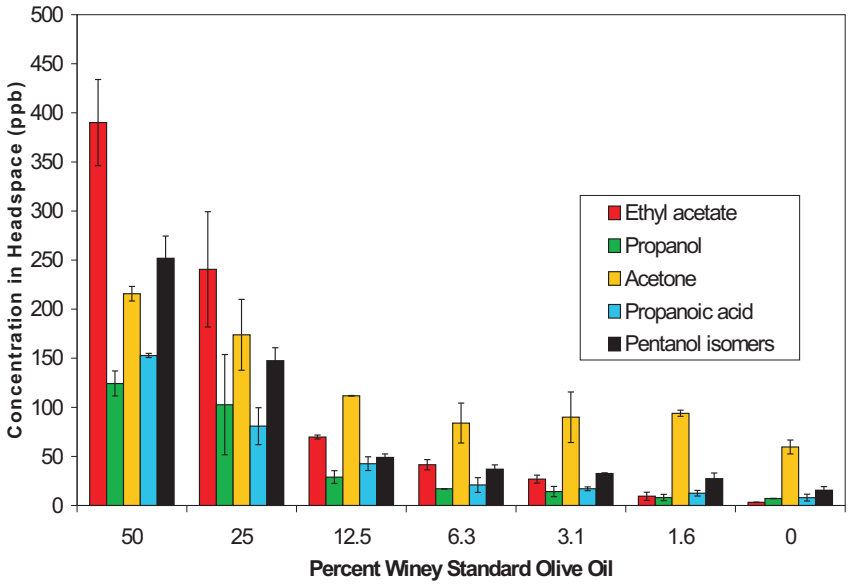



**Figure 5.10.** Concentrations of five VOCs measured above olive oil samples with different proportions of winey standard olive oil mixed for the group of tasters. All five of these VOCs are present at noticeably higher concentrations in winey olive

oil than in fresh olive oil. Values are means ± standard deviations of duplicate measurements.

The greatest value in this experiment was the opportunity to determine the concentrations of VOCs unique to winey olive oil at the sensory detection threshold for the defect. Provided these VOCs are produced in the same relative abundances in all winey olive oil as they were present in the defective standard oil, this will allow the detection of the winey defect in olive oil by SIFT-MS. However, as all VOCs displayed a strong correlation with the proportion of winey standard olive oil added, only those known to be attributed to winey olive oil should be considered. They are ethanol, ethyl acetate and acetic acid. The relationship found between ethanol and the winey defect is shown in figure 5.11, while the relationships found for all three VOCs are given in table 5.5.

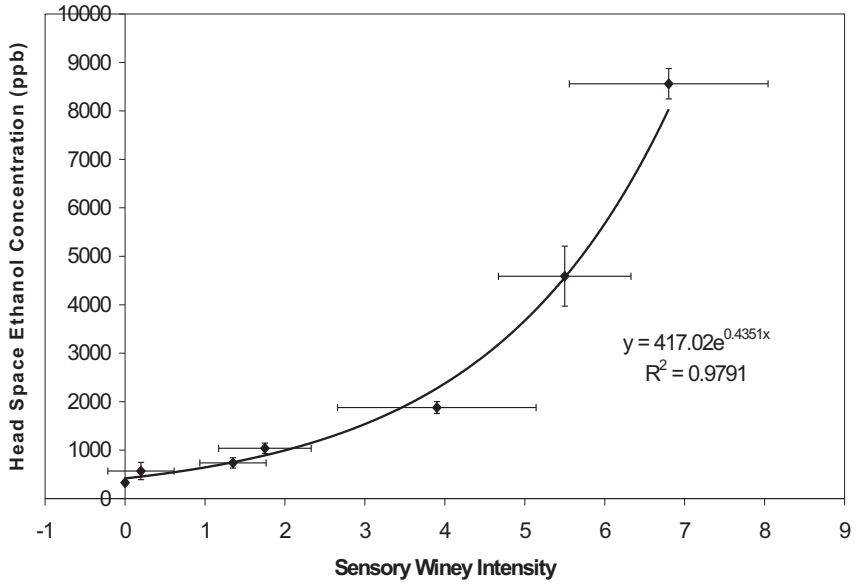


**Figure 5.11.** The relationship found between the head space ethanol concentration and the intensity of the winey sensory defect. Values for ethanol concentrations are means ± standard deviations of duplicate measurements, values for sensory intensity are medians ± robust standard deviations(*9*).

**Table 5.5.** Equations of best fit and $R^2$ values for head space concentrations of three VOCs with the sensory evaluation of the winey defect.

| VOC | Curve fitted | $R^2$ |
|---|---|---|
| Ethanol | $y = 417e^{0.435x}$ | 0.979 |
| Ethyl acetate | $y = 7.58e^{0.613x}$ | 0.919 |
| Acetic acid | $y = 38.9e^{0.128x}$ | 0.879 |

The pre-exponential factors listed in table 5.5 are the y intercepts of the best fit lines, and hence are approximations of the VOC concentrations at the sensory detection limit of the winey defect.

Unfortunately, due to the lack of more samples exhibiting this defect and the inexperience of the tasting panel (which would not have improved significantly had the investigation been pursued further), the study was concluded at this point.

A slight alteration to the method used in the winey defect study involves the addition of the defective olive oil to a fresh oil, to simulate the emergence of the defective VOC products in the presence of VOCs from fresh oil(*12;13*). Fresh oil was not used in the present study for several reasons. The most important is that the standard IOOC method for determining the sensory threshold of a sensory panel for a given defect uses purified, tasteless oil similar to the refined oil used in this study(*14*). The IOOC method is well established and widely used by sensory panels

and for that reason was followed as closely as possible. It was hoped that data from other similar samples tasted by the ONZ sensory panel at Hort Research would be made available, however this information is collected by the IOOC for accreditation purposes and is not released, even to the sensory panels themselves.

## 5.4 Musty Olive Oil

The musty olive oil standard, like that for rancid olive oil, showed elevated intensities for the peaks at m/z = 57, 88 and 99 with the $NO^+$ precursor corresponding to propanal, acetone and hexanal compared with fresh oil (figure 5.12). These are the same VOCs found at elevated concentrations in the rancid defect and they are also present in many oils identified as defective by sensory panels. Rancidity can set in quickly under appropriate conditions, where the oil is heated, exposed to light or broken down by other processes. There were no distinctive volatile compounds identified by SIFT-MS analysis that were present in musty olive oil which set it apart from other types of defective olive oil. In fact, all VOCs were found to be at very low concentrations.

The most common VOCs linked to the musty defect by olive oil researchers are 1-octen-3-one and the corresponding alcohol, 1-octen-3-ol. Both of these VOCs have an earthy, mouldy aroma, and are typical metabolites of fungal organisms(*15*). The head space of musty olive oil was subjected to a SIM scan monitoring the masses of expected products from reaction of 1-octen-3-one and 1-octen-3-ol with $H_3O^+$ (m/z = 127 and m/z = 129 respectively) and $NO^+$ (m/z = 156 and m/z = 127 respectively). This scan is shown in figure 5.13. No significant response was

observed for either of these expected products, suggesting they are present at very low concentrations in the head space of the standard oil obtained.

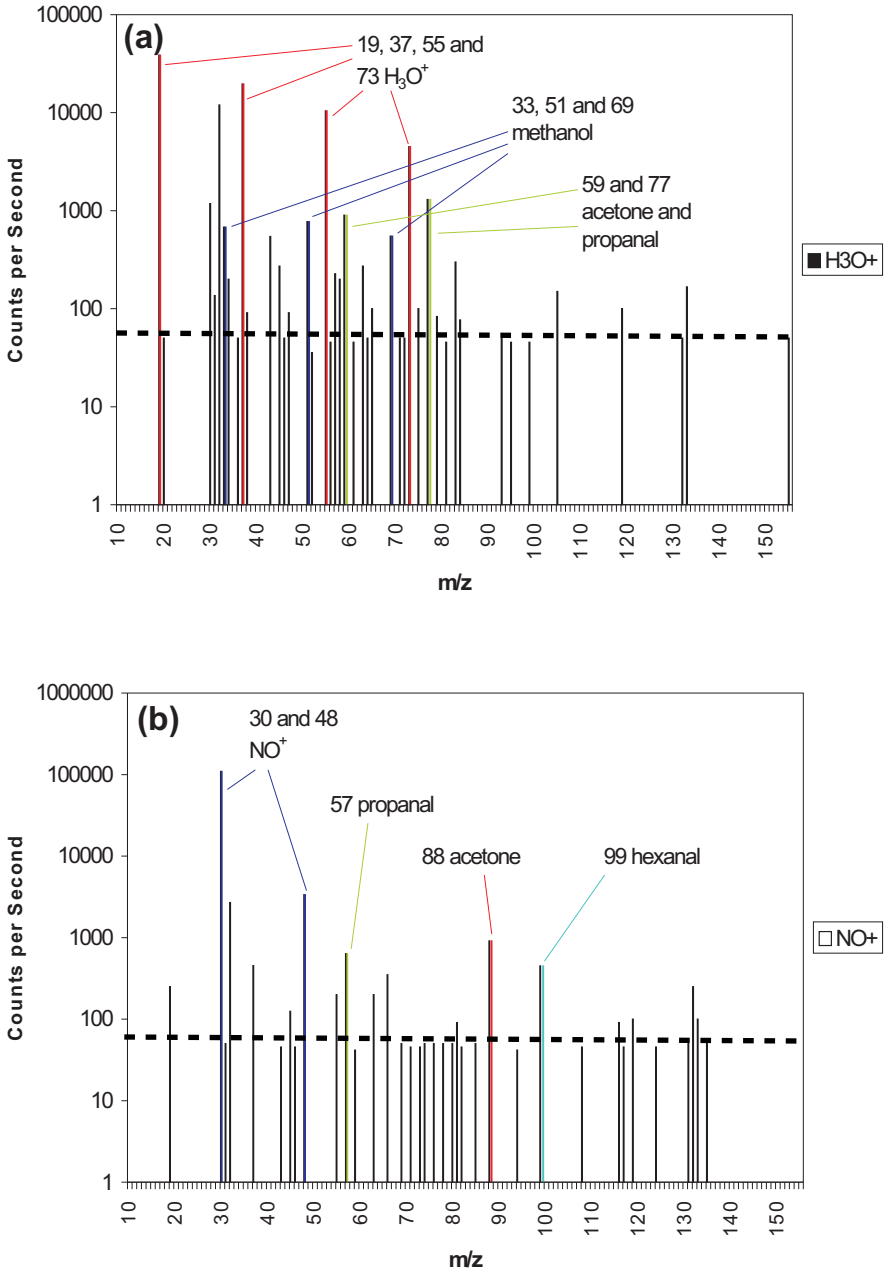



**Figure 5.12.** SIFT-MS mass scans of head space above musty standard olive oil with the **(a)** $H_3O^+$ reagent ion and **(b)** $NO^+$ reagent ion.

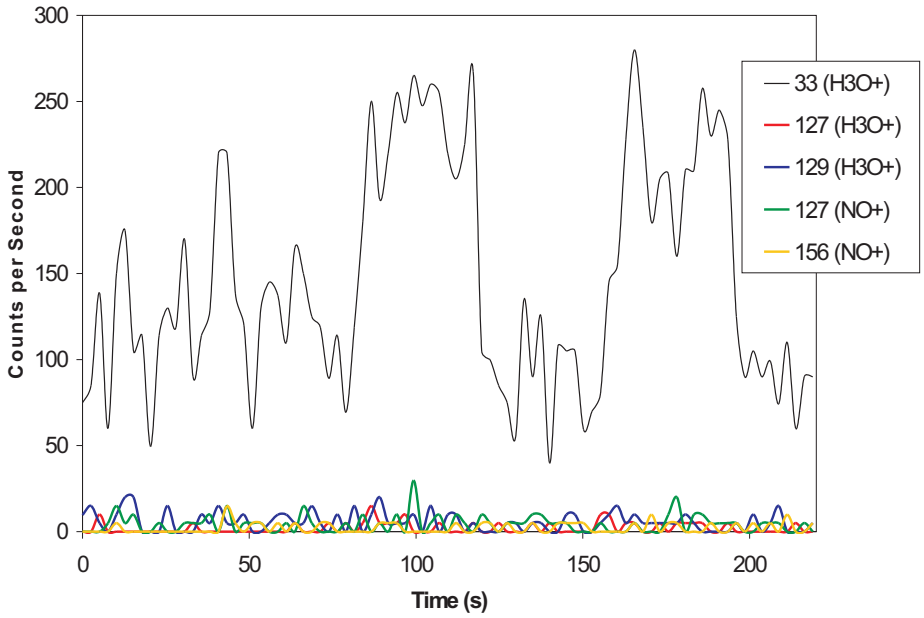



**Figure 5.13.** Selected m/z values monitored during the course of a SIM scan of musty olive oil head space. The oil was analysed twice, the times of sampling are indicated by the rise in response of methanol at m/z = 33 (the $CH_3OH.H^+$ ion). No significant change was observed for any of the other m/z values, which correspond to the VOCs of interest.

No simple method of identifying the musty defect in olive oil was discovered from this study. One consolation is that to the best of our knowledge, the musty defect does not occur in New Zealand olive oil. For this defect to occur, olives must be stored in large piles or sacks for several days between harvesting and processing. Processors are sufficiently abundant (or conversely growers are sufficiently scarce) so as to avoid the conditions which bring about the musty defect. As a consequence, the only source for less common defects such as musty is

the IOOC in Spain, where standard defects intended for panellist training are available.

## 5.5 Fusty Olive Oil

Initial SIFT-MS analysis of fusty olive oil head space showed higher concentrations of both methanol and acetone than typical fresh olive oils, and at lower concentration, ethyl acetate was also observed (figure 5.14).

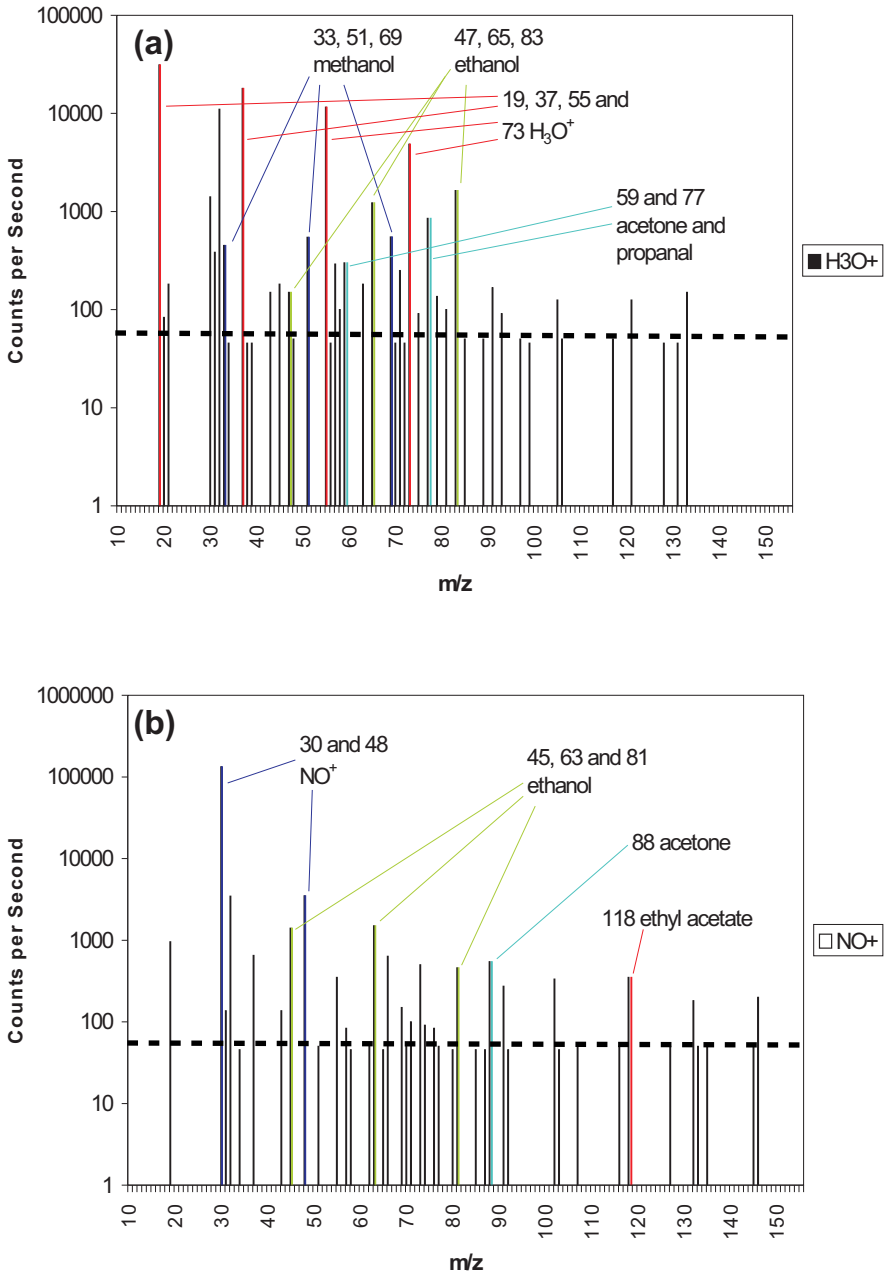



**Figure 5.14.** SIFT-MS mass scans of head space above fusty standard olive oil with the **(a)** $H_3O^+$ reagent ion and **(b)** $NO^+$ reagent ion.

Angerosa et. al.(*16*) found that 3-methyl-1-butanol was produced by the anaerobic fermentation which causes the fusty defect. 3-methyl-1-butanol can be

detected but not differentiated from other pentanol isomers using the $H_3O^+$ or $NO^+$ reagent ions in SIFT-MS. The inability to distinguish between these isomers is not considered a serious shortcoming, as very low concentrations of these VOCs have been observed in all fresh oils and the majority of oils displaying different defects. Hence, the formation of other pentanol isomers in olive oil is considered to be by only minor reaction channels.

Two mixtures containing different proportions of fusty standard oil were presented to the group of local tasters. Both mixtures were evaluated as having a fustiness intensity of five and containing no other defects. SIFT-MS analysis showed no positive correlation between the head space concentration of pentanol isomers and the proportion of the fusty standard oil added. Little can be done with these results, as only a limited volume of oil was available (which restricted the number of samples able to be prepared) and little correlation was observed between the composition of the samples, the tasting data and the VOC profiles.

The only significant correlation observed was that between ethyl acetate and the proportion of fusty standard olive oil in the sample (figure 5.15). As shown in figure 5.14.(b), the fusty defect shares ethyl acetate as a product with the winey defect. It is not known whether the fusty and winey defects are able to be distinguished from each other by SIFT-MS analysis, as only one sample of each defect was able to be obtained and both appeared very similar. It is possible that both defects were present in each standard, as the development of these defects are favoured under very similar conditions. Again, as with the musty defect, the fusty

and winey defects are very uncommon in New Zealand and examples of oils displaying them were too few to allow thorough investigation.

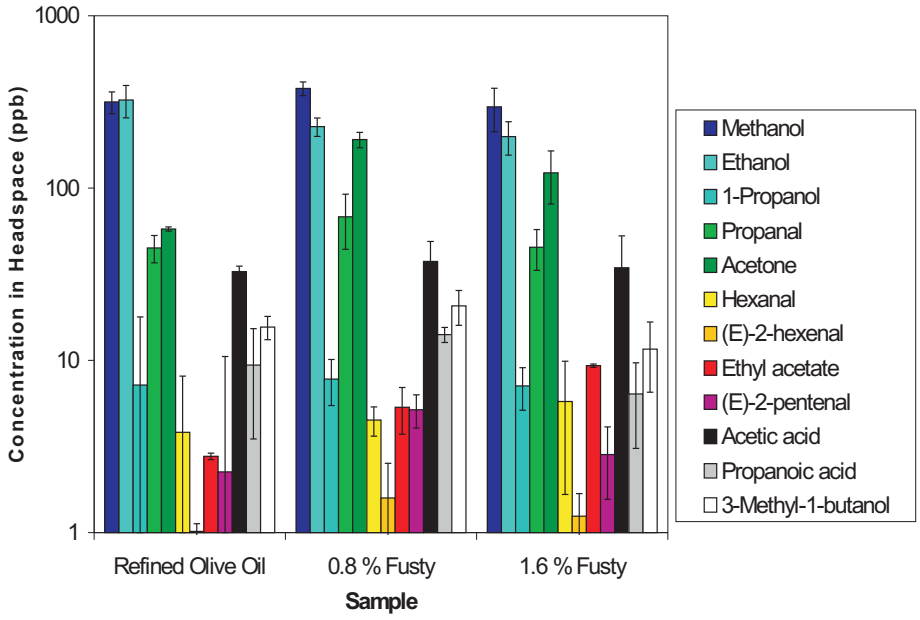



**Figure 5.15.** VOC concentrations in the head space above refined olive oil with different proportions (percent by volume) of a fusty standard oil added. Values are means ± standard deviations of triplicate measurements.

## 5.6 Muddy Olive Oil

The olive oil which had been contaminated with muddy sediment produced the most interesting spectra of any olive oil obtained, which were also among the most difficult to interpret. They are shown in figure 5.16. There are many high mass peaks present, mostly due to increased concentrations of carboxylic acids (such as acetic acid) and a very large increase in the concentration of ethanol present compared with any other oil analysed.

The peak at m/z = 121 and part of that at m/z = 139 in the $H_3O^+$ spectrum (figure 5.16.(a)) can be attributed to acetic acid, being the acetic acid dimer cluster and its water cluster respectively. The high concentration of ethanol also caused the formation of several extra cluster products at m/z = 111 and 139, corresponding to the ethanol dimer water cluster and the ethanol trimer cluster. For dimer and trimer clusters of ethanol to occur, the concentration of ethanol in the head space must be in the mid-to-high ppm range. The mass peaks at m/z = 107 and 125 are the first and second water clusters of ethyl acetate respectively. The peaks at m/z = 135 and 149 are both clusters of protonated propanoic acid, the first with acetic acid and the second with unprotonated propanoic acid.

A number of new products observed in the $NO^+$ spectrum (figure 5.16.(b)) also did not appear in any other oils. The product at m/z = 118 is a primary product of ethyl acetate. It was included simply to draw attention to its high intensity. The m/z = 105 peak may be attributed to the water cluster of one or several isomers of pentanal which were discussed in the previous chapter. The peak at m/z = 91 may be a water cluster of an isomer of butanol or the ethanol dimer cluster. There is an associated water cluster at m/z = 109 which may result from either butanol or ethanol products. The m/z = 73 peak is large enough to be the primary butanol product, however the corresponding peak in the $H_3O^+$ spectrum (m/z = 75) is of low intensity. Subsequent experiments have demonstrated that ethanol can form a dimer cluster at m/z = 91 with the $NO^+$ reagent ion. If butanol were present at high concentration, the primary $H_3O^+$ product would form a dimer cluster, giving a

product at m/z = 151. The corresponding peak is not present in figure 5.16(a), so butanol is unlikely to be present.

Large peaks were observed at m/z = 132 and 146 in the $NO^+$ spectrum. These represent a progression from m/z = 118 with a difference of m/z = 14 between each value. This is often characteristic of a VOC series, as the $CH_2$ group has m/z = 14. In this case it is likely to represent esters, as m/z = 118 is the main $NO^+$ product of ethyl acetate. The acetic acid and propanoic acid clusters observed in the $H_3O^+$ spectrum at m/z = 121 and 135 may have some contribution from these esters, which are isomers of propyl acetate and propyl propanoate. The ester isomers cannot be identified from inspection of these spectra, and it is difficult to discern if their presence is reinforced by the $H_3O^+$ spectrum. The primary $H_3O^+$ peaks at m/z = 103 and 117 are small, however the water cluster peaks which are expected (as these larger esters should behave much like ethyl acetate does and cluster readily with water) coincide with those of other compounds. The single water clusters for the two sizes of ester would be found at m/z = 121 and 135. These peaks coincide with the acetic acid dimer cluster and the acetic acid-propanoic acid cluster respectively. The second water cluster products would appear at m/z = 139 and 153. The ethanol trimer cluster is found at m/z = 139, however there is a small peak at m/z = 153 that is not attributable to any other compound. The peak is so small that it may easily be declared noise.

As mentioned in the discussion on winey olive oil, the presence of some VOCs at high concentration can lower the accuracy of quantitative measurements made by SIFT-MS. The solution to this is to lower the concentrations of these

VOCs in the flow tube by decreasing the sample flow rate. This also has the effect of decreasing the sensitivity and increasing the detection limit of the corresponding measurements. For mass scans such as those performed in this chapter, quantitation was not important, only identification. Therefore a high sample flow rate was permitted to provide more intense product ion peaks. This had the effect of producing more cluster ion peaks which would not otherwise be observed. The new product ion peaks observed in these samples which were not seen in chapter 4 were not added to the SIM scan method used for quantitation, as they would increase the scan time dramatically and are not expected to be present when acceptable (i.e. negligible) levels of reagent ion depletion are observed.

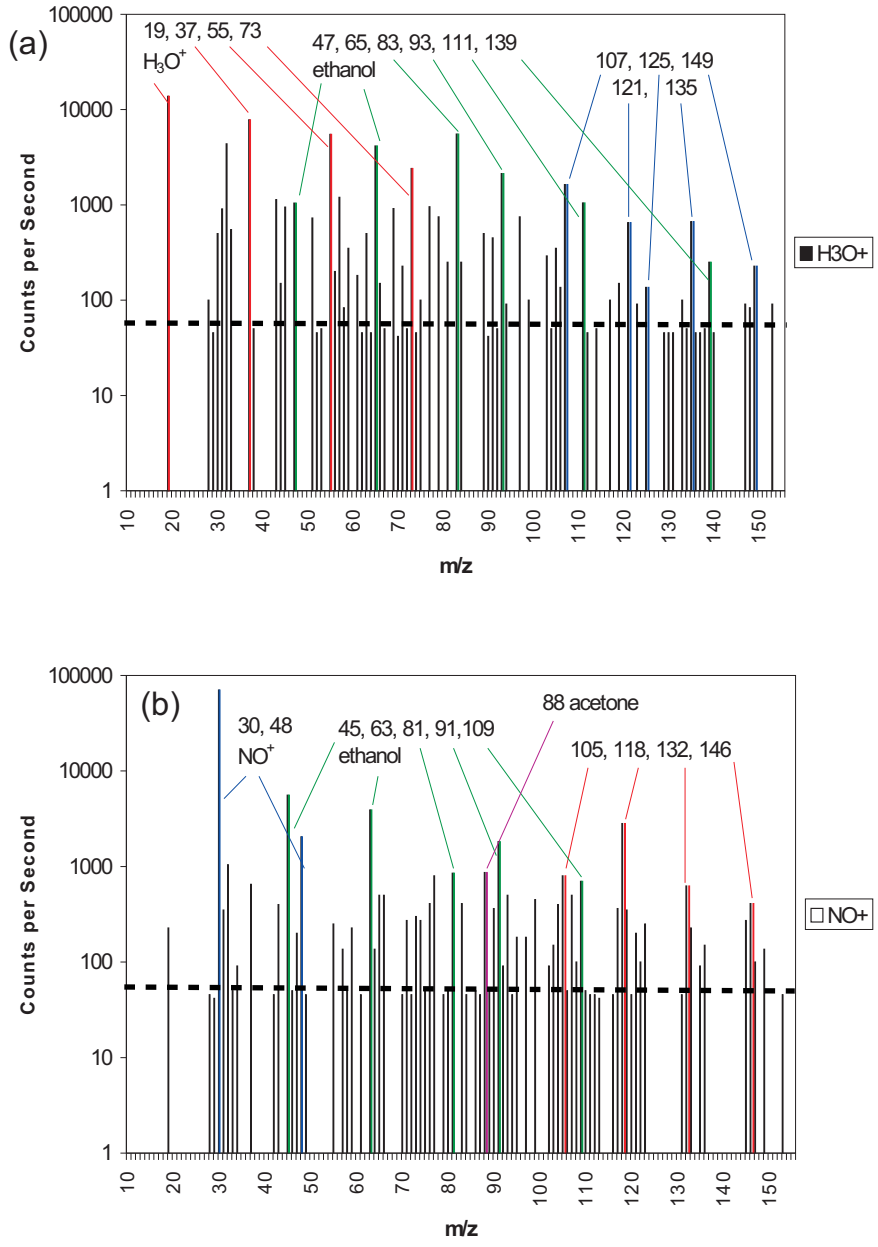



**Figure 5.16.** SIFT-MS mass scans of head space above muddy standard olive oil with the **(a)** $H_3O^+$ reagent ion and **(b)** $NO^+$ reagent ion.

The muddy defect is another which is due to fermentation; not of olives, but of oil during storage. Due to the constraints of time and oil availability, a detailed

study involving oil samples displaying the muddy defect was not performed, although it presents an interesting challenge. More oil would be necessary for an investigation of the muddy defect than for the rancid defect, as more sensory data would be required. Sensory analysis is the only established test for the analysis of the muddy defect, and the ONZ sensory panel requires 250 mL of oil for each sample. Therefore, for a study similar to that conducted for the rancid defect, 2.5 litres of each of three oils would be needed. This volume of oil would provide 14 analyses by SIFT-MS and 7 sensory analyses per oil. The opportunity for an undertaking of this scale was not available during the present research, but the potential offered by SIFT-MS for a simple diagnostic test for the muddy defect is high. The muddy defect study is recommended as future work, as it may help to identify this defect – one of the few defects that has appeared in New Zealand olive oil.

## 5.7 References


(1) Belitz, H.-D.; Grosch, W. *Food Chemistry*, 2nd English ed.; Springer-Verlag: Berlin, Germany, 1999.

(2) Medina, I.; Satué-Gracia, M. T.; German, J. B.; Frankel, E. N. Comparison of Natural Polyphenol Antioxidants from Extra Virgin Olive Oil with Synthetic Antioxidants in Tuna Lipids during Thermal Oxidation. *J. Agric. Food Chem.* **1999**, *47*, 4873-4879.

(3) Frankel, E. N.; Satué-Gracia, T.; Meyer, A. S.; German, J. B. Oxidative Stability of Fish and Algae Oils Containing Long-Chain Polyunsaturated Fatty Acids in Bulk and in Oil-in-Water Emulsions. *J. Agric. Food Chem.* **2002**, *50*, 2094-2099.

(4) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 639-660.

(5) Procida, G.; Giorno, A.; Cichelli, A.; Conte, L. S. Study of Volatile Compounds of Defective Virgin Olive Oils and Sensory Evaluation: A Chemometric Approach. *J. Sci. Food Agric.* **2005**, *85*, 2175-2183.

(6) Solinas, M.; Angerosa, F.; Cucurachi, A. Connessione tra Prodotti di Neoformazione Ossidativa delle Sostanze Grasse e Insorgenza del Difetto di Rancidità all'Esame Organolettico. Nota 1. *Riv. Soc. It. Sci. Alim.* **1985**, *14*, 361-368.

(7) Solinas, M.; Angerosa, F.; Cucurachi, A. Connessione tra Prodotti di Neoformazione Ossidativa delle Sostanze Grasse e Insorgenza del Difetto di Rancidità all'Esame Organolettico. Nota 2. Determinazione Quantitativa. *Riv. Ital. Sost. Grasse* **1987**, *44*, 137-145.

(8) Španel, P.; Smith, D. Selected Ion Flow Tube Studies of the Reactions of $H_3O^+$, $NO^+$ and $O_2^+$ with Several Aromatic and Aliphatic Hydrocarbons. *Int. J. Mass Spectrom.* **1998**, *181*, 1-10.

(9) International Olive Oil Council COI/T.20/Doc. no. 15/Rev. 1 Sensory Analysis of Olive Oil Method: Organoleptic Assessment of Virgin Olive Oil. http://www.internationaloliveoil.org/downloads/orga6.pdf (accessed Feb 13, 2007)

(10) Martens, H. A.; Dardenne, P. Validation and Verification of Regression in Small Data Sets. *Chemom. Int. Lab. Sys.* **1998**, *44*, 99-121.

(11) Gomez-Alonso, S.; Salvador, M.; Fregapane, G. Evolution of the Oxidation Process in Olive Oil Triacylglycerol under Accelerated Storage Conditions (40-60°C). *J. Amer. Oil Chem. Soc.* **2004**, *81*, 177-184.

(12) Aparicio, R.; Rocha, S. M.; Delgadillo, I.; Morales, M. T. Detection of Rancid Defect in Virgin Olive Oil by the Electronic Nose. *J. Agric. Food Chem.* **2000**, *48*, 853-860.

(13) García-Gonzalez, D. L.; Aparicio, R. Detection of Vinegary Defect in Virgin Olive Oils by Metal Oxide Sensors. *J. Agric. Food Chem.* **2002**, *50*, 1809-1814.

(14) International Olive Oil Council COI/T.20/Doc. no. 14/Rev. 1 Guide for the Selection, Training and Monitoring of Skilled Virgin Olive Oil Tasters. http://www.internationaloliveoil.org/downloads/orga5eng.pdf (accessed Jun 5, 2007)

(15) Morales, M.; Luna, G.; Aparicio, R. Comparative Study of Virgin Olive Oil Sensory Defects. *Food Chem.* **2005**, *91*, 293-301.
(16) Angerosa, F.; Di Giacinto, L.; Solinas, M. Influenza dello Stoccaggio in Massa delle Olive sull'Aroma degli Oli di Risulta: Valutazione del Difetto di "Riscaldo" Mediante Analisi HPLC e GLC dei Componenti Volatili. *Riv. Merceol.* **1990**, *29*, 275-294.

# Chapter 6

## Olive Oil Sensory Attributes

### 6.1 Introduction

The present chapter reports the results of two studies. Both involved SIFT-MS VOC analysis and tasting of commercially available olive oil samples. The first study included 33 oils which were tasted by a group of local tasters. All were familiar with olive oil and most were habitual olive oil users, however this group had had no formal training in the tasting of olive oil. The second study included 40 oils analysed by the Olives New Zealand tasting panel during the 2006 Extra Virgin Olive Oil certification program run by Olives New Zealand. These data were generated by a panel of well trained and regularly tested judges, yet were also more difficult to obtain. Principal Component Analysis (PCA) was first applied to both data sets to explore the relationships between the different sensory attributes and also the relationships between the VOCs. Partial Least-Squares (PLS) regression was then applied to the results from the Olives New Zealand panel to determine the relationships between the sensory data and the instrumental (VOC) data with the aim of predicting the sensory attributes using the VOC concentrations measured for the oil samples.

### 6.2 Method

#### 6.2.1 Sensory Analysis

Sensory analysis was performed by two different groups: the Olives New Zealand sensory panel based at HortResearch in Auckland, New Zealand and a less

formal collection of local tasters based at Syft Technologies in Christchurch, New Zealand. The ONZ panel is accredited by the IOOC for the assessment of olive oil quality characteristics. Olive oils were assessed using the standard assessment sheet on a computer, where the strength of an attribute is rated as a distance from the left (0 attribute strength) of a line 10 cm in length (10 cm denotes an attribute strength of 10). The attributes were divided into positive (fruity, bitter, pungent) and negative (fusty, musty, winey, muddy sediment, metallic, rancid or other).

The less formal collection of tasting judges evaluated olive oils on a different scale from the ONZ panel. Here oils are rated out of 100, with each attribute allotted a maximum value based on its perceived importance to overall oil acceptability. This rating system is not as well regulated and contains more opportunity for bias than does the standard assessment system. However, it also provides more detailed sensory information.

### 6.2.2 SIFT-MS Analysis

$5 \pm 0.1$ mL of olive oil was placed in a 500 mL clear glass bottle (Schott Glass, Mainz, Germany), full capacity $632 \pm 0.5$ mL, with a silicone rubber septum cap and left to stand for 20 minutes at $24 \pm 0.2$ °C to allow a good partition of volatiles between the liquid and the gas phases. All SIFT-MS analyses were performed using the LDI#1 SIFT-MS instrument at Syft Technologies. The bottle headspace was sampled via a needle into the SIFT-MS flow tube through a heated stainless steel capillary at a controlled rate of 1.88 Torr L $s^{-1}$ (148 mL $min^{-1}$ at standard pressure). This was introduced to the flow tube which contained He

flowing at 57.6 ± 0.1 Torr L $s^{-1}$ and Ar flowing at 9.3 ± 0.2 Torr L $s^{-1}$, with a total internal tube pressure of 0.80 ± 0.05 Torr (106 ± 7 Pa).

### 6.2.3 Statistical Analysis

The statistical methods used to analyse the present data were Principal Component Analysis (PCA) and Partial Least Squares (PLS) regression. PCA and PLS were performed using Matlab 7, Release 13. PCA was applied via the method outlined by Reyment and Jöreskog(*1*) and PLS was applied via the method outlined by Haaland and Thomas(*2*). Both methods are described in appendix A. The code, written as part of this research, is given in appendix C, section C.2.

## 6.3 Results

### 6.3.1 Principal Component Analysis

Data from the ONZ sensory panel were difficult to obtain, therefore the less well controlled data using the local tasters were generated in the hope of obtaining the largest possible set of detailed tasting data. Thirty three olive oils were analysed by the local tasting group, while the results of forty olive oils evaluated by the ONZ panel were also obtained. The data from each panel were analysed separately, as the two sets of results were not able to be combined. Each sensory panel used different criteria to measure different aspects of the oil against different standards. The locally generated data served as important preliminary information and provided insight into the preferences of habitual, yet untrained olive oil consumers. The data from this informal group of tasters were not reliable enough to be employed for quantitative purposes such as method development, as there was no formal training,

no selection criteria and the correct conditions for tasting olive oil were not implemented. The results from these sessions remain interesting, however, and their analysis is described first.

PCA output matrices and vectors have been given distinctive names throughout this chapter. Along with those outlined in section A.2 (**L**, **C** and **F** for eigenvalue vectors, loading and score matrices respectively), those matrices and vectors concerning sensory data contain the label **SENS**, while those concerning instrumental VOC concentration data contain the label **INST**. The method of scaling is identified by either a **G** or **Z** (the input matrices referring to mean-centred and standardised data respectively) and a **1** or **2** indicate the data being analysed (**1** for locally generated tasting data and **2** for data from the Olives New Zealand sensory panel).

***Locally Generated Sensory Data***

Several of the sensory attributes measured were found to be very highly correlated with each other (table 6.1), such as olfactory greenness and gustatory greenness. Greenness is an aroma or taste similar to grass, leaves or unripe fruit. There is a minimum recommended ratio of samples to attributes for PCA (somewhere between 5:1 and 10:1(*3*)) to obtain the most reliable results. Data pre-treatment was carried out to increase this ratio in the hope of obtaining the best possible results from the PCA method. The 'olive fruitiness' attributes ('olive fruitiness' is the taste and smell of ripe, healthy olives), evaluated both nasally and retronasally, are difficult to define and hence were excluded, as without sufficient training no tasters were expected to be able to adequately identify or quantify this

attribute. It was also highly correlated with several other attributes, most notably the harmony/balance attribute. The olive fruitiness attribute would be expected to introduce solely noise, with all of its important information already described by the harmony/balance attribute.

**Table 6.1.** Correlation matrix for the sensory attributes. Attributes with nasal and retronasal components were added together to leave the eight attributes listed. Several remaining attributes are strongly correlated, presenting the opportunity for further reduction in the number of attributes included in the PCA. Correlation matrices are symmetrical about their diagonal elements, which are all one, as the correlation of any attribute with itself is one.

| | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* |
|---|---|---|---|---|---|---|---|---|
| Olive fruitiness (*1*) | 1 | 0.89 | 0.89 | 0.94 | 0.07 | 0.4 | 0.67 | 0.85 |
| Complexity (*2*) | 0.89 | 1 | 0.82 | 0.94 | 0.11 | 0.44 | 0.66 | 0.81 |
| Greenness (*3*) | 0.89 | 0.82 | 1 | 0.88 | -0.05 | 0.43 | 0.65 | 0.91 |
| Harmony/balance (*4*) | 0.94 | 0.94 | 0.88 | 1 | 0.11 | 0.42 | 0.61 | 0.87 |
| Sweetness (*5*) | 0.07 | 0.11 | -0.05 | 0.11 | 1 | -0.02 | -0.08 | 0 |
| Bitterness (*6*) | 0.4 | 0.44 | 0.43 | 0.42 | -0.02 | 1 | 0.65 | 0.47 |
| Pungency (*7*) | 0.67 | 0.66 | 0.65 | 0.61 | -0.08 | 0.65 | 1 | 0.75 |
| Persistence (*8*) | 0.85 | 0.81 | 0.91 | 0.87 | 0 | 0.47 | 0.75 | 1 |

It was decided that any attributes displaying correlations of greater than 0.9 with each other would be combined to form a single attribute. As a result, olfactory greenness and gustatory greenness were added together to produce an attribute simply called greenness. Other attributes with sufficiently high correlations yet which were not so obviously connected were also added together, such as greenness with persistence and harmony/balance with complexity. However, there were also correlations between very different attributes which were slightly lower than the

arbitrary correlation limit of 0.9 for adding attributes together, yet were still significant. The responses for greenness/persistence and those for harmony/balance/complexity are highly correlated (0.89), yet the two sets of attributes are conceptually different. Only fresh, intense oils from unripe fruit would be expected to score highly for greenness and persistence, while a delicate, mature, mild oil may still display harmony, balance and complexity. With a correlation of this magnitude the attributes are most likely measuring the same variation. However, it was decided to perform PCA without combining these attributes, as PCA is designed to identify correlations and combine attributes into principal components where necessary. The pre-treatment that was applied up to this point was performed with some certainty that no important information was being lost. However this pre-treatment is much more clumsy than the PCA algorithm, and as the samples-to-attributes ratio was now 6.5:1, it was no longer necessary to continue combining attributes.

Different scaling regimes were used for the data sets employed in this research, as the measurements are the results of different systems. The SIFT-MS variables are VOCs which vary markedly in their concentrations. Methanol and ethanol in particular are consistently present in the sample head space at concentrations much greater than those of the majority of other compounds. However, due to the different sensitivities of human sensory organs to different compounds, the absolute concentration of a VOC does not necessarily reflect its importance in determining oil quality as judged by a sensory panel. For example, acetic acid is able to be detected by humans at 42 ppm, while the threshold

concentration for sensory detection of methanol is more than 800 times higher, at 34 000 ppm(*4*). Therefore even at much lower concentrations acetic acid may display a greater effect on the perceived oil quality than the much more prevalent methanol. Hence, the SIFT-MS head space results were standardised to produce a **Z** matrix (see section A.1.1) before PCA was performed. A more rigorous method of scaling may be to divide the concentration of each compound by its respective sensory detection threshold, yet this presents problems. Sensory detection thresholds (the concentration at which a compound is detectable to humans by the appropriate sense, in this case taste or smell) are not available for all relevant VOCs. Additionally, each compound has a nasal and retronasal (sensed from outside the body and from within the mouth respectively) detection threshold, and these two values are not correlated. Olive oil tasting utilises both nasal and retronasal detection, making a choice between the two very difficult. Adjusting the data with these considerations in mind may not produce values which are any more relevant than unscaled data. Hence a simple standardisation of instrumental results was employed.

The tasting data were generated by rating each sensory attribute on a predetermined scale – the preliminary tasting data weighted each attribute differently, while the official panel evaluated all attributes from zero to ten – zero being the lowest and ten being the highest strength of each sensory attribute. The official panel data should not require standardisation. The question of whether the preliminary data requires standardisation does not have an obvious answer, therefore it will be investigated.

The preliminary data from the local panel do not rate all attributes on an equal scale, however the attribute values are scaled by their perceived importance to the overall oil flavour. This would suggest that the data not be standardised before PCA is performed. However, as PCA has a tendency to display only the most important relationships and hide the rest when covariances (as opposed to correlations) are considered, PCA was also performed on the standardised data matrix so that relationships involving attributes with low perceived importance may also be detected.

PCA was first carried out on the unstandardised deviate score (**G**) matrix. The eigenvalue vector (**LSENSG1**) and loadings matrix (**CSENSG1**) are shown in table 6.2. LSENSG1 shows how many PCs are required to describe the data by representing the importance of each PC by the size of the corresponding eigenvalue. As suspected, those sensory attributes with the highest absolute values (which often also possess the highest variance) completely dominate. The trace of the covariance matrix (and hence also the sum of the eigenvalues) is 99.45. The first PC, therefore, represents almost 98 % of the variation in the data and the others may be discarded. The combined harmony/balance and complexity attribute has a loading of 1.00 on the first and only significant PC, which shows that the only significant relationship detected is completely dominated by this one attribute. The combined greenness and persistence attribute, being closely correlated with harmony/balance and complexity, also has a high loading (0.90) on the first PC for mean-centred data. The other attributes, predictably, have lower loadings (pungency 0.64, bitterness 0.43 and sweetness 0.10).

**Table 6.2.** Eigenvalues (**LSENSG1**) and loadings (**CSENSG1**) for mean-centred tasting data obtained from the tasting of 33 olive oils performed by the local sensory panel. The size of each eigenvalue represents the amount of variation in the data accounted for by the corresponding principal component, and therefore how important that principal component is for describing the data. The values of the loadings on each principal component show which attributes are closely correlated with each other.

| LSENSG1 | | CSENSG1 | |
|---|---|---|---|
| PC 1 | 97.34 | Attribute | PC 1 |
| PC 2 | 1.37 | Harmony/balance & complexity | 1.00 |
| PC 3 | 0.44 | Greenness & persistence | 0.90 |
| PC 4 | 0.18 | Sweetness | 0.10 |
| PC 5 | 0.12 | Bitterness | 0.43 |
| | | Pungency | 0.64 |

The eigenvalue vector (**LSENSZ1**) for the data derived from the standardised deviate score (**Z**) matrix is shown in table 6.3, along with its scree plot in figure 6.1 (see section A.2 for an explanation of these terms). The slope of the plot does not allow a simple determination of the correct number of principal components to use, however three seems reasonable. Another criterion (which is only of use for standardised data) which recommends using only those principal components with eigenvalues greater than 1 suggests the use of two principal components. As all sensory attributes in standardised data have the same variance and the sum of the eigenvalues is equal to the number of sensory attributes, any eigenvalue less than one does not contribute any more in the way of insight into the structure of the data than the original attributes did. This rule of thumb suggests that

all PCs corresponding to eigenvalues less than one should be discarded as noise. “Noise” in PCA refers to variation in the data which is not due to correlations between the measured variables. The goal of PCA is to describe correlations in the data. Labelling the remaining variation as “noise” shows that no interesting information can be obtained from it. Two principal components were thus chosen to describe the data.

**Table 6.3.** Relevant output data from PCA for the standardised tasting data. More than one pattern of variation was highlighted (as more than one principal component was found to be significant), therefore more information is gained from the standardised data than from the unstandardised data. **LSENSZ1** describes the amount of variation in the original data accounted for by each PC. **CSENSZ1** shows the correlation between each sensory attribute and each significant PC. **FSENSZ1** shows the value of each oil as measured by each significant PC. The values in **FSENSZ1** are obtained by multiplying the original data by their corresponding loading values from **CSENSZ1** as in equation A.7.

| **LSENSZ1** | | **CSENSZ1** | | | |
|---|---|---|---|---|---|
| PC 1 | 2.91 | | | PC 1 | PC 2 |
| PC 2 | 1.13 | Harmony/balance & complexity | | 0.872 | -0.303 |
| PC 3 | 0.586 | Greenness & persistence | | 0.915 | -0.166 |
| PC 4 | 0.284 | Sweetness | | -0.0783 | -0.927 |
| PC 5 | 0.0913 | Bitterness | | 0.725 | 0.382 |
| | | Pungency | | 0.883 | 0.0758 |
| | | | | | |
| **FSENSZ1** | | | | | |
| | PC 1 | PC 2 | | PC 1 | PC 2 |
| Oil 1 | 0.954 | 0.104 | Oil 18 | -0.262 | -1.91 |
| Oil 2 | -1.84 | -0.18 | Oil 19 | 0.236 | -0.813 |
| Oil 3 | -0.91 | -0.0791 | Oil 20 | 0.537 | 2.79 |
| Oil 4 | 1.21 | 0.799 | Oil 21 | -0.259 | -0.462 |
| Oil 5 | 0.329 | 0.804 | Oil 22 | 0.505 | -0.227 |
| Oil 6 | -1.6 | 1.12 | Oil 23 | 0.789 | -0.795 |
| Oil 7 | -0.615 | 0.776 | Oil 24 | 0.321 | -0.813 |
| Oil 8 | 0.863 | 0.104 | Oil 25 | 0.86 | 1.11 |
| Oil 9 | 0.113 | 0.583 | Oil 26 | 0.912 | 1.23 |
| Oil 10 | -2.48 | 0.434 | Oil 27 | 0.767 | 0.743 |
| Oil 11 | 0.422 | 0.341 | Oil 28 | 1.04 | -0.124 |
| Oil 12 | 0.471 | -0.271 | Oil 29 | -0.2 | -0.964 |
| Oil 13 | -1.92 | 1.48 | Oil 30 | -0.714 | -0.428 |
| Oil 14 | 0.581 | -1.18 | Oil 31 | -0.311 | -1.27 |
| Oil 15 | 0.203 | -1.8 | Oil 32 | 1.42 | 0.556 |
| Oil 16 | -0.336 | -0.962 | Oil 33 | -1.75 | -0.151 |
| Oil 17 | 0.675 | -0.547 | | | |

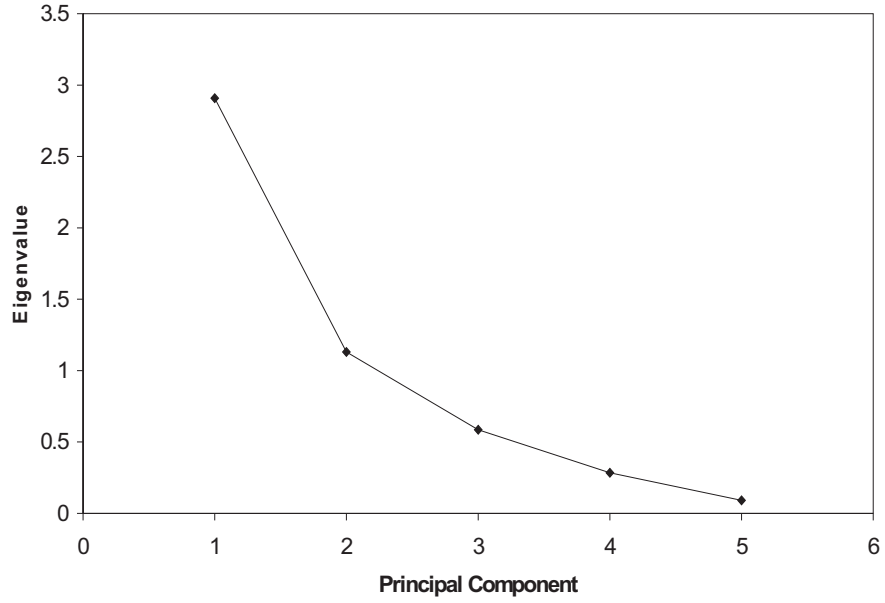



**Figure 6.1.** Scree plot for eigenvalues derived from PCA of standardised tasting data. There is no clear change in slope to provide a basis from which to exclude subsequent PCs. Considering only those eigenvectors from **LSENSZ1** greater than one as being significant, two PCs are included.

The loadings matrix (**CSENSZ1**) is shown in table 6.3, and gives some interesting information on the PCs chosen. The correlations between the sensory attributes and the first PC are close to one for all attributes except sweetness. On PC 2 sweetness alone shows a high correlation. This suggests that all attributes apart from sweetness are well correlated with each other and are found to a greater degree in all high quality oils, and to a lesser degree in all low quality oils. This result may appear obvious, however mild-tasting, ripe oils would not be expected to score high values for greenness or pungency, yet still may be awarded high values for harmony/balance or complexity. This result suggests that the local group of

tasters would award low values for all attributes to any oil that did not display strong greenness and pungency. Sweetness is the exception as it was observed in oils with any degree of greenness or pungency and at any stage of oxidation, regardless of the strengths of any other attributes.

The PC scores matrix (**FSENSZ1**) is included in table 6.3, however the same information is contained in figure 6.2 and is much more intuitive when displayed in this form. **FSENSZ1** may be used to determine which oils are similar to each other, if any oils form a cluster apart from the rest, and to identify any interesting structure in the data. Those oils to the right in figure 6.2 were awarded high values for all attributes apart from sweetness, while those to the bottom were awarded high values for sweetness. The present data do not show any surprising features, however the maximum value is noticeably less than 100, and there is a group of five oils off to the left which were considered by the tasters to be worse than the other oils. These are oils 2, 6, 10, 13 and 33. Four of these five oils were store-bought imported olive oils and the other was an oil produced locally from frosted fruit. Figure 6.2 demonstrates well the ability of PCA to organise and efficiently describe multivariate data.

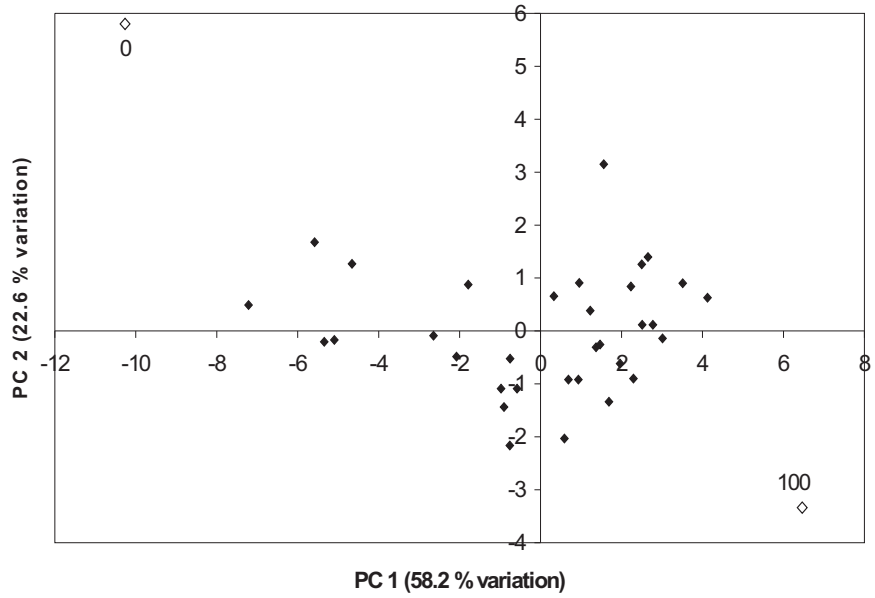



**Figure 6.2.** Principal component scores for tasting data scaled by their eigenvalues to account for the variation covered by each principal component. PC 1 represents all attributes except sweetness, PC 2 represents mostly the sweetness attribute. The open diamonds marked with 0 and 100 correspond to hypothetical oils which were awarded 0 and 100 % respectively of the maximum values in order to gauge the values awarded to those oils analysed. The hypothetical oils were projected onto the PC space using the pattern (**A**) matrix and eigenvectors as shown in equation A.7.

The eigenvalue vector for the instrumental VOC measurements (**LINSTZ1**) is shown in table 6.4. This provides the amount of variation in the data accounted for by each PC. Its scree plot (figure 6.3) suggests six PCs, while there are only three eigenvalues above one. Considering the cumulative variance accounted for by each PC and all previous PCs, at least 90 % gives a good approximation, while over 95 % is desirable. Six PCs contain almost 95 % of the data, the same number

suggested by the scree plot. Six PCs were chosen. The large number of PCs prohibits plotting of the PC scores and leaves the loadings matrix as the only source of information for interpretation.

**Table 6.4.** Eigenvalues for standardised instrumental data (VOC concentrations). The cumulative percent column gives the sum of the corresponding eigenvalue and all previous eigenvalues as a percentage of the sum of all ten eigenvalues. With the inclusion of six PCs, close to 95 % of the variation in the data is explained. This only excludes four PCs, so it is not possible to reduce these data to the same degree as it was for the corresponding sensory data.

| **LINSTZ1** | | Cumulative % |
|---|---|---|
| PC 1 | 4.14 | 41.4 |
| PC 2 | 1.95 | 60.9 |
| PC 3 | 1.32 | 74.1 |
| PC 4 | 0.958 | 83.6 |
| PC 5 | 0.71 | 90.7 |
| PC 6 | 0.389 | 94.6 |
| PC 7 | 0.21 | 96.7 |
| PC 8 | 0.182 | 98.5 |
| PC 9 | 0.113 | 99.7 |
| PC 10 | 0.0324 | 100 |

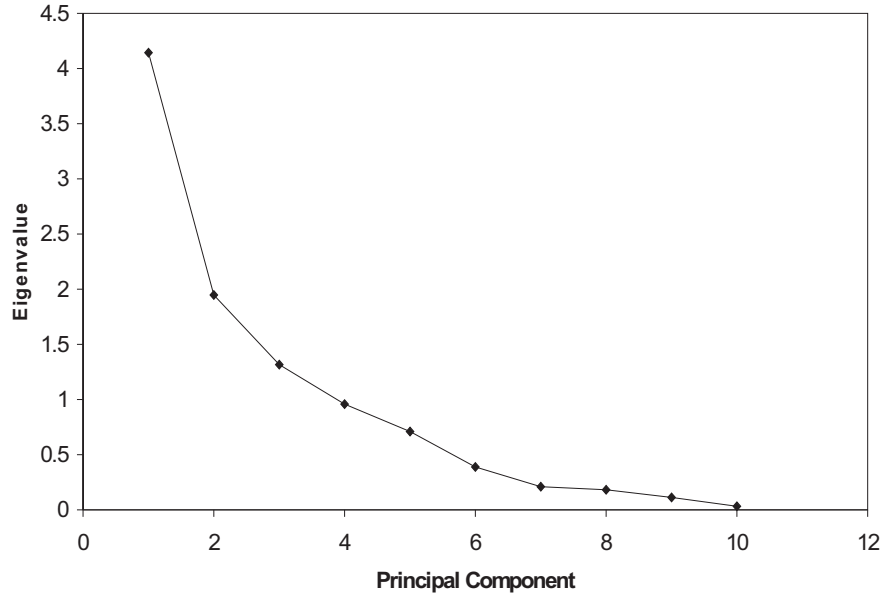



**Figure 6.3.** Scree plot of eigenvalues for standardised instrumental data. The slope does not become close to zero until six PCs are included. This suggests that the last four PCs are only accounting for noise, and no useful descriptive information is lost by excluding them.

The loadings matrix (**CINSTZ1**, table 6.5) is more complicated than the loadings matrix for the sensory measurements as more VOCs were measured than sensory attributes. Methanol, ethanol, hexanal, ethyl acetate and propanoic acid all display correlations above 0.7 with the first PC. Propanoic acid has a correlation of 0.89, which is very significant. So the major feature of the data is the correlation between these five compounds and it is their combined variation which is described by the first PC. The second PC has high correlations with propanal and (E)-2-hexenal, although each has a different sign. It is not unreasonable to link the first PC with the degree of fermentation, as propanoic acid, ethyl acetate and

ethanol (all common fermentation products) are well represented. The second PC is strongly linked to the degree of oxidation of the oil. An effort was made to include oils at different stages of oxidation, with some oils very unpleasant due to oxidation. This is reflected in the second PC, where propanal (a major oxidation product identified in section 5.2) displays a positive correlation and (E)-2-hexenal displays a negative correlation. This result is interesting, as it shows that oxidised olive oils have lower concentrations of (E)-2-hexenal than fresh oils.

**Table 6.5.** Correlations of each VOC with each significant PC. Only the first two PCs have significant correlations which may be interpreted, linking PC 1 with the degree of fermentation (high correlations with methanol, ethanol, hexanal, ethyl acetate and propanoic acid) and PC 2 with the degree of oxidation (high correlations with propanal and acetone and a high negative correlation with (E)-2-hexenal).

| CINSTZ1 | | | | | | |
|---|---|---|---|---|---|---|
| VOC | PC 1 | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 |
| Methanol | 0.773 | -0.159 | -0.417 | -0.289 | 0.263 | 0.144 |
| Ethanol | 0.74 | 0.002 | 0.14 | -0.0483 | -0.609 | -0.0982 |
| Propanal | 0.3 | 0.738 | 0.073 | 0.532 | 0.193 | 0.0755 |
| Acetone | 0.408 | 0.689 | -0.471 | 0.254 | -0.0921 | -0.0812 |
| Hexanal | 0.797 | -0.278 | 0.272 | 0.148 | 0.279 | 0.213 |
| (Z)-3-hexenal | 0.569 | -0.184 | -0.741 | -0.0417 | -0.0712 | -0.0504 |
| (E)-2-hexenal | -0.128 | -0.742 | -0.216 | 0.482 | 0.167 | -0.311 |
| Ethyl acetate | 0.77 | 0.116 | 0.416 | -0.0418 | 0.159 | -0.379 |
| (E)-2-pentenal | 0.616 | -0.472 | 0.158 | 0.424 | -0.269 | 0.231 |
| Propanoic acid | 0.892 | 0.0581 | 0.167 | -0.297 | 0.123 | -0.0669 |

Considering the group of five oils identified from **FSENSZ1** in figure 6.2 as being of below average sensory quality (oils 2, 6, 10, 13 and 33), the first two columns of **FINSTZ1** (shown in table 6.6) may give some insight into their

characteristic VOC profiles. On PC 1, which most likely measures the degree of fermentation, only oils 13 and 33 have high scores. However, the five oils have the five highest scores on PC 2, which measures the degree of oxidation. This result suggests that the degree of oil oxidation, while not changing an oil's VOC profile as much as fermentation does (as oxidation is represented by PC 2 in **CINSTZ1** and fermentation by PC 1, where PC 1 represents twice as much variation as PC 2 as judged from **LINSTZ1**), has a greater effect on the oil's sensory properties than fermentation. The correlation between the degree of oxidation and the sensory score does not exist for any other oils, however the results obtained for these five oils support the conclusion from the previous chapter that propanal and acetone are the major volatile oxidation products of olive oil, as these two VOCs have the highest loadings on PC 2 in the **CINSTZ1** loadings matrix.

**Table 6.6**. The first two columns (corresponding to the first two PCs) of the PC scores for the VOC concentration data (**FINSTZ1**). The PC 1 scores represent the degree of fermentation experienced by the oils, while the PC 2 scores represent the degree of oxidation.

| FINSTZ1 | | | | | |
|---|---|---|---|---|---|
| | PC 1 | PC 2 | | PC 1 | PC 2 |
| Oil 1 | -0.741 | -0.459 | Oil 18 | -0.405 | -0.222 |
| Oil 2 | -0.0918 | 1.08 | Oil 19 | 0.0574 | -0.324 |
| Oil 3 | -0.525 | 0.562 | Oil 20 | -0.589 | -0.359 |
| Oil 4 | -0.526 | -0.225 | Oil 21 | 2.3 | -0.264 |
| Oil 5 | -0.603 | -1.42 | Oil 22 | 0.0433 | -0.955 |
| Oil 6 | -0.0055 | 1.49 | Oil 23 | 1.61 | 0.207 |
| Oil 7 | -0.0411 | 0.423 | Oil 24 | 2.68 | -1.13 |
| Oil 8 | -0.429 | -1.29 | Oil 25 | -1.19 | 0.235 |
| Oil 9 | -0.534 | -1.38 | Oil 26 | 0.0027 | -0.737 |
| Oil 10 | -0.0548 | 3.32 | Oil 27 | -0.377 | 0.532 |
| Oil 11 | -0.846 | -0.0045 | Oil 28 | -0.179 | -0.413 |
| Oil 12 | 0.381 | -0.0983 | Oil 29 | -0.497 | -0.218 |
| Oil 13 | 1.01 | 1.88 | Oil 30 | -0.823 | 0.218 |
| Oil 14 | -0.653 | 0.099 | Oil 31 | 2.55 | -0.847 |
| Oil 15 | -0.785 | 0.067 | Oil 32 | 0.254 | -0.618 |
| Oil 16 | -1 | -0.458 | Oil 33 | 0.854 | 1.61 |
| Oil 17 | -0.852 | -0.31 | | | |

***Olives New Zealand Panel Data***

Next, the results from the 40 oils evaluated by the ONZ panel were subjected to PCA. These oils were evaluated for the strengths of three important sensory attributes: fruitiness, bitterness and pungency. As discussed earlier, the tasting data here were not standardised, while the instrumental data were.

The eigenvalue (**LSENSG2**) vector obtained from PCA of the ONZ panel results is shown in table 6.7. As there are only three PCs, plotting the eigenvalues does not give sufficient additional information. The third PC is considered to contain mostly noise, so two PCs appropriately represent these data.

**Table 6.7.** Eigenvalues from PCA for mean-centred sensory data from the ONZ panel. Two PCs were deemed sufficient to represent the data, as the eigenvalue corresponding to PC 3 is noticeably smaller than those of the other two PCs.

| | **LSENSG2** | Cumulative % |
|---|---|---|
| PC 1 | 1.21 | 60.2 |
| PC 2 | 0.57 | 88.4 |
| PC 3 | 0.233 | 100 |

The PC loadings (**CSENSG2**) matrix is shown in table 6.8. Bitterness and pungency both display significant correlations with the first PC, while fruitiness does not. Fruitiness alone is correlated significantly with the second PC, however. This result shows that the strengths of the bitterness and pungency attributes are related in many olive oils, whereas the strength of the fruitiness attribute is independent.

**Table 6.8.** PC loadings (**CSENSG2**) and scores (**FSENSG2**) matrices. Bitterness and pungency are described by the first PC in **CSENSG2**, while the second describes the fruitiness of the oils. The elements of **FSENSG2**, scaled by their respective eigenvalues, are plotted in figure 6.4.

| **CSENSG2** | | |
|---|---|---|
| | PC 1 | PC 2 |
| Fruitiness | 0.464 | 0.885 |
| Bitterness | 0.881 | -0.183 |
| Pungency | 0.888 | -0.249 |

| **FSENSG2** | | | | | |
|---|---|---|---|---|---|
| | PC 1 | PC 2 | | PC 1 | PC 2 |
| Oil 1 | 0.363 | -0.0536 | Oil 21 | 1.24 | -0.645 |
| Oil 2 | -0.903 | 0.602 | Oil 22 | -0.392 | 0.486 |
| Oil 3 | 0.213 | 1.02 | Oil 23 | 0.22 | 0.424 |
| Oil 4 | -1.72 | -0.23 | Oil 24 | 0.579 | -0.367 |
| Oil 5 | 0.601 | 0.403 | Oil 25 | -1.39 | -0.829 |
| Oil 6 | -0.815 | -0.146 | Oil 26 | -0.569 | 1.98 |
| Oil 7 | -0.149 | -0.784 | Oil 27 | -0.337 | 1.04 |
| Oil 8 | -0.192 | -1.38 | Oil 28 | 0.505 | -1.09 |
| Oil 9 | 1.74 | -1.3 | Oil 29 | -1.52 | 0.336 |
| Oil 10 | -1.9 | -3.04 | Oil 30 | 1.3 | 0.154 |
| Oil 11 | -0.333 | 0.174 | Oil 31 | 0.0625 | 0.384 |
| Oil 12 | 0.0366 | 0.81 | Oil 32 | -0.613 | -0.364 |
| Oil 13 | 1.29 | 0.77 | Oil 33 | -0.708 | 0.692 |
| Oil 14 | -0.352 | 0.0056 | Oil 34 | 0.213 | 1.02 |
| Oil 15 | 1.3 | 0.154 | Oil 35 | -0.15 | 0.0738 |
| Oil 16 | -1.92 | -0.0642 | Oil 36 | 0.334 | 0.669 |
| Oil 17 | -0.63 | -0.257 | Oil 37 | 0.427 | 0.756 |
| Oil 18 | 1.03 | -1.54 | Oil 38 | 0.657 | 1.81 |
| Oil 19 | -0.669 | -0.349 | Oil 39 | -0.18 | -0.908 |
| Oil 20 | 2.7 | -1.79 | Oil 40 | 0.63 | 1.39 |

The PC score (**FSENSG2**) matrix, scaled by the appropriate eigenvalues, is plotted in figure 6.4. There is no useful structure and no discernable clusters of oils from inspection of this chart.

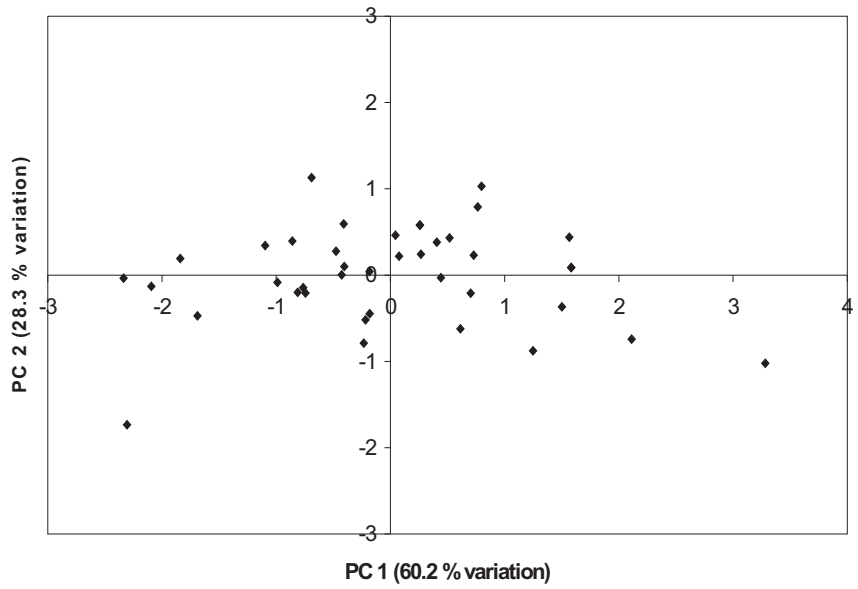



**Figure 6.4.** Scaled PC scores (scaled **FSENSG2**) for sensory data from the ONZ panel. PC 1 represents bitterness and pungency, PC 2 represents fruitiness. Those oils plotted further to the right display more bitterness and pungency, while those further to the top display more fruitiness.

The eigenvalue vector for the instrumental variables (**LINSTZ2**) is shown in table 6.9. Its scree plot (figure 6.5) suggests either four or eight PCs, depending on whether the slope of the curve from PCs five to eight is considered to be close enough to zero for these PCs to be discarded. The eigenvalues measure the variance accounted for by their respective PCs and are ordered from greatest to least. At the point where the eigenvalues become similar to each other (the curve on the scree plot has a slope of approximately zero), all important variation is considered to have been accounted for and all subsequent PCs are discarded as measuring only noise. Considering the cumulative variance accounted for by the PCs suggests that

at least six should be used, but no more than eight. Six PCs were selected. As with the INSTZ1 data set, the high number of PCs makes the loadings matrix the best opportunity for interpretation.

**Table 6.9.** Eigenvalues and the cumulative variance accounted for by them from the SIFT-MS VOC data collected for 40 olive oils. Six PCs were chosen to represent these data, retaining more than 88 % of the original variation.

| | **LINSTZ2** | Cumulative % |
|---|---|---|
| PC 1 | 4.43 | 34.1 |
| PC 2 | 2.69 | 54.8 |
| PC 3 | 1.67 | 67.6 |
| PC 4 | 1.29 | 77.5 |
| PC 5 | 0.787 | 83.6 |
| PC 6 | 0.626 | 88.4 |
| PC 7 | 0.483 | 92.1 |
| PC 8 | 0.394 | 95.1 |
| PC 9 | 0.233 | 96.9 |
| PC 10 | 0.16 | 98.1 |
| PC 11 | 0.121 | 99.1 |
| PC 12 | 0.0877 | 99.7 |
| PC 13 | 0.0326 | 100 |

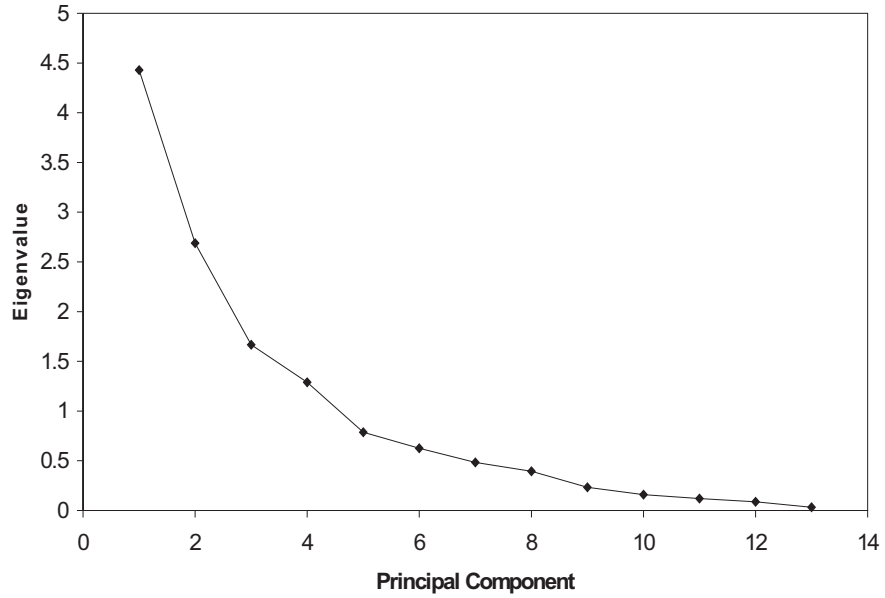



**Figure 6.5.** Scree plot of eigenvalues from **LINSTZ2**. Either four or eight PCs are suggested from this plot, depending on whether the slope of the points from five to eight PCs is considered to be significantly greater than zero. This is a subjective choice, with six PCs being selected to represent these data.

The PC loadings (**CINSTZ2**) matrix is shown in table 6.10. The highest loading on the first PC is for hexanal, at 0.83. All other loadings are between 0.5 and 0.7 apart from ethanol, (E)-2-hexenal and acetic acid. The first PC measures some source of variation where all VOCs except ethanol, (E)-2-hexenal and acetic acid are correlated. Similar situations are presented by the second, and in fact all PCs, albeit with lower correlations and slightly wider variations between them. No useful information about trends among these oils is obvious from the loadings matrix, which suggests there is little difference between them.

**Table 6.10.** PC loadings of VOCs on the six significant PCs found from analysis of 40 olive oils.

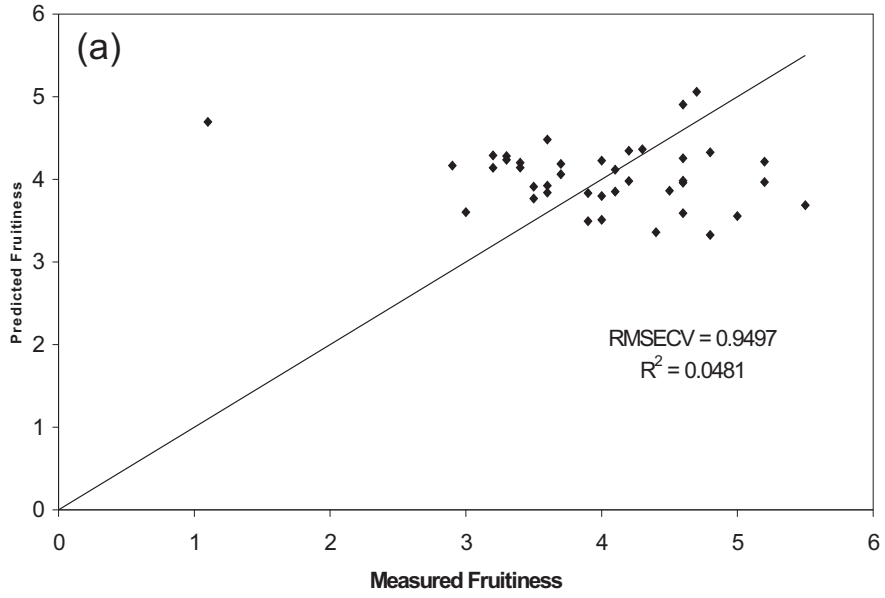
| CINSTZ2 | | | | | | |
|---|---|---|---|---|---|---|
| VOC | PC 1 | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 |
| Methanol | 0.501 | 0.562 | 0.457 | -0.222 | 0.0415 | -0.253 |
| Ethanol | 0.407 | 0.559 | -0.633 | 0.157 | 0.0225 | -0.109 |
| Propanol | 0.624 | 0.535 | -0.498 | 0.0299 | 0.0457 | 0.0932 |
| Propanal | 0.581 | -0.552 | -0.125 | 0.0678 | -0.32 | -0.0786 |
| Acetone | 0.6 | -0.402 | 0.138 | -0.229 | 0.206 | 0.565 |
| Hexanal | 0.825 | -0.0971 | -0.41 | -0.0773 | 0.207 | 0.107 |
| (Z)-3-hexenal | 0.616 | 0.107 | 0.438 | -0.449 | -0.26 | 0.0569 |
| (E)-2-hexenal | 0.108 | -0.775 | -0.0704 | -0.298 | 0.188 | -0.346 |
| Ethyl acetate | 0.658 | -0.275 | 0.261 | 0.566 | -0.174 | 0.0323 |
| (E)-2-pentenal | 0.664 | -0.23 | -0.317 | -0.346 | -0.381 | -0.152 |
| Acetic acid | 0.43 | -0.538 | 0.011 | 0.215 | 0.461 | -0.205 |
| Propanoic acid | 0.635 | 0.0964 | 0.342 | 0.588 | -0.0853 | -0.0401 |
| Butanol isomers | 0.628 | 0.51 | 0.352 | -0.174 | 0.313 | -0.116 |

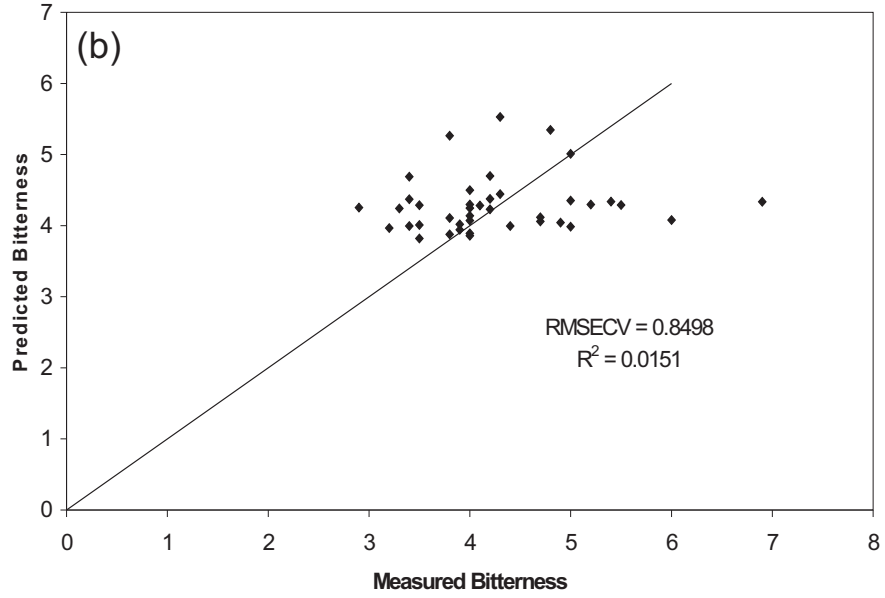
### 6.3.2 Partial Least-Squares Regression

PLS regression was performed on the data produced by the ONZ panel. Each attribute was treated individually and cross-validation was performed using the leave-one-out method. The data from the local group of tasters was not considered reliable enough to warrant the use of quantitative regression methods.

The predicted attribute values from leave-one-out cross-validation are plotted against the measured values in figure 6.6. RMSECV (a measure of prediction errors from cross-validation) and $R^2$ values are given in the respective plots. All prediction models were of rank one, that is one latent variable provided the best prediction according to the RMSECV and $R^2$ values. No attribute was well predicted from the data available. As these were the only data available, the study must be regarded as unsuccessful. The focus of this investigation now shifts to

discovering the reason or reasons behind the unsuitability of the data and suggesting future actions which may avoid this for subsequent investigation.



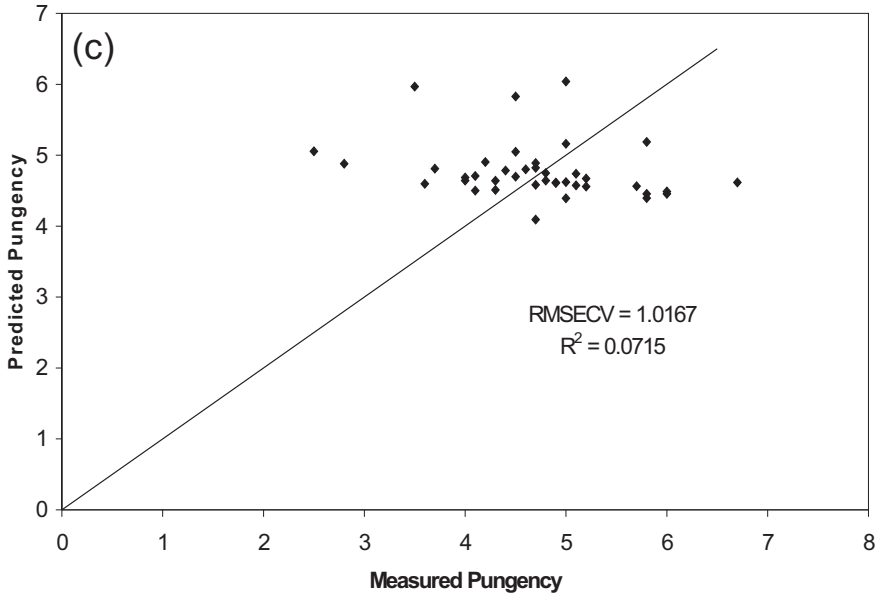
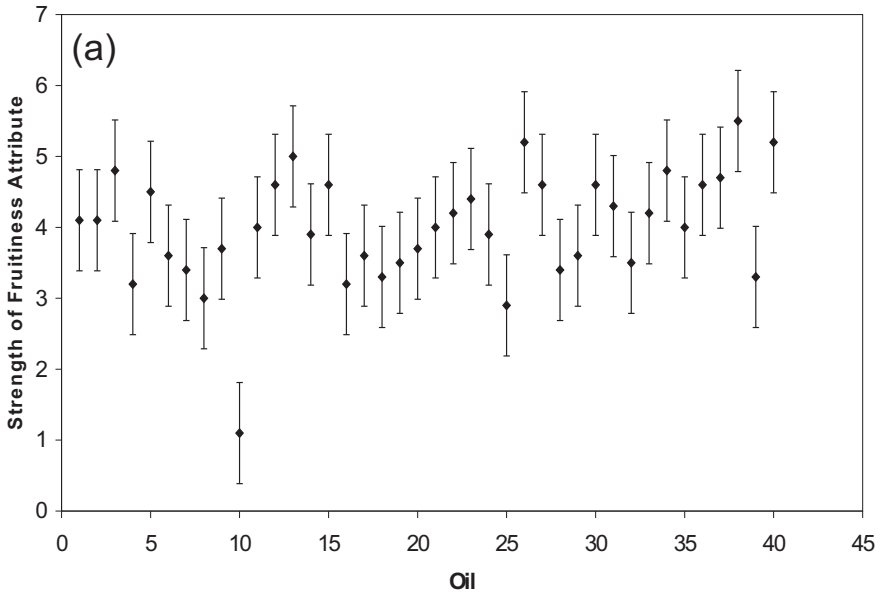
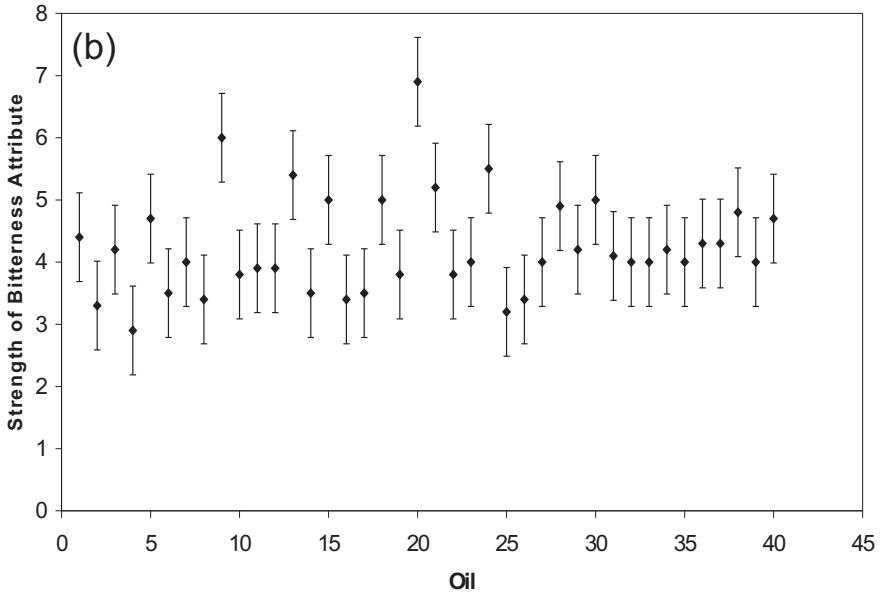



**Figure 6.6.** Plots of measured versus predicted values for the three sensory attributes **(a)** fruitiness, **(b)** bitterness and **(c)** pungency. The solid black line in each plot represents the function y = x, along which all data would lie if well predicted. All prediction models were of rank one.

The PCA results may help to identify why the sensory attributes were so poorly predicted. The sensory results (SENSG2) do not highlight anything unusual, however the instrumental results (INSTZ2) suggested that most of the oils were essentially identical. If there were very little variation between the oils, that would explain why almost all oils for each attribute were predicted to have the same intensity. The tasting data (the raw SENSG2 data, as obtained from the Olives New Zealand sensory panel) are plotted in figure 6.7. The variation between the intensities of sensory attributes, in accordance with the instrumental PCA results (INSTZ2), is very low compared with the expected error of each measurement. As

there was no indication given of the errors associated with the sensory measurements, they were taken to be the average value of the robust standard deviation determined for rancidity from the study reported in section 5.2. All measurements in the rancidity study were found to have an almost constant robust standard deviation of ± 0.71.



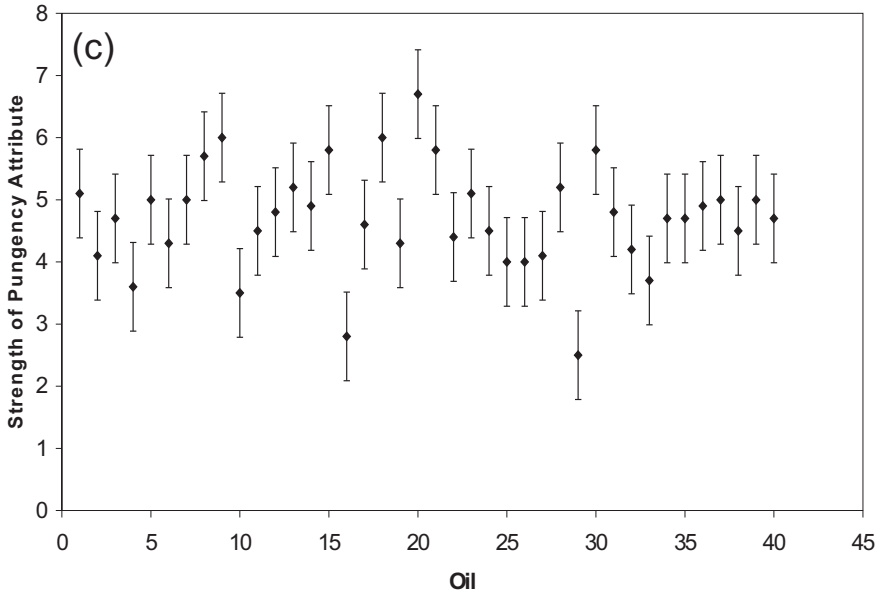



**Figure 6.7.** Sensory data for the 40 olive oils obtained from the ONZ sensory panel showing the evaluated intensities of **(a)** the fruitiness, **(b)** the bitterness and **(c)** the pungency attributes. Almost all oils were awarded very similar values for each attribute.

For the PLS regression technique to identify important trends and correlations in a data set, there must be sufficient variation in the data. This is not the case in the present data, so the intensities of sensory attributes for all oils were predicted to be the means of the respective attributes.

In order for future studies correlating sensory and instrumental data to be successful, at least the same number of oils as used in this study are necessary. One hundred oils would be a more appropriate number, if this could be arranged. The values for all sensory attributes and VOC concentrations must span the entire range of values likely to be encountered in the future, and should be approximately

normally distributed over that range. The measured values for the VOCs and sensory attributes used in the present study did not cover a wide enough range of values to sufficiently expose the correlations between the sensory and instrumental data.

## 6.4 Summary

The data obtained from the local olive oil tasters contained high correlations between every attribute except sweetness. A large range of oils was tasted by this group, and it is difficult to believe that all but one of the attributes were truly correlated to that degree. The lack of proper training and facilities was highlighted as a possible reason behind the high correlations. The corresponding instrumental data contained information on the degree of oxidation and possibly fermentation of the oils, judged solely on the VOCs correlated with each PC and knowledge of the circumstances under which different VOCs are produced.

The data obtained from the ONZ sensory panel showed that the intensities of bitterness and pungency were correlated in the oils analysed, and the intensity of fruitiness was independent of the other attributes. The instrumental data did not expose any relationships which could be easily interpreted.

The attempt to predict the intensities of sensory attributes from VOC concentrations using PLS regression on the ONZ sensory panel data was unsuccessful. This was believed to be due to the narrow range of values spanned by the data. The PLS regression method could not characterise the variation in the data because there was too little variation in the data to characterise.

## 6.5 References

(1) Reyment, R.; Jöreskog, K. *Applied Factor Analysis in the Natural Sciences*; Cambridge University Press: New York, NY, USA, 1993.
(2) Haaland, D.; Thomas, E. Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem.* **1988**, *60*, 1193-1202.
(3) Hair, J. J.; Anderson, R.; Tatham, R.; Black, W. *Multivariate Data Analysis*, 4th ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 1995.
(4) Abraham, M. H.; Kumarsingh, R.; Cometto-Muniz, J. E.; Cain, W. S. An Algorithm for Nasal Pungency Thresholds in Man. *Arch. Toxicol.* **1998**, *72*, 227-232.

# Chapter 7

## Antioxidant Introduction

### 7.1 Introduction

One of the reasons olive oil is highly valued is its high content of antioxidants. These naturally occurring compounds react with highly reactive radical species to form more stable products, thus protecting many important biological molecules from damage. Other vegetable oils also contain antioxidants and the concentrations of antioxidant compounds can vary widely between different olive oils. It is not sufficient to claim a high antioxidant content for an oil; it must be measured. Before the antioxidant content of an oil can be measured, we must define what an antioxidant is. A precise definition is difficult(*1*), however a useful definition which applies to the present research is:

"any chemical species which may be added to a system at relatively low concentration to prevent or slow the reaction of other species with radicals".

The antioxidants found in olive oil are described in section 1.2. The present chapter describes important radical species and some tests developed to assess the antioxidant potential of samples, called 'antioxidant capacity'.

### 7.2 Biologically Important Radical Species

Halliwell and Gutteridge(*2*) list the following as the most important biological radicals: transition metals, hydroxyl radicals ($OH^{\bullet}$), superoxide radicals ($O_2^{\bullet}$), peroxyl and alkoxyl radicals ($ROO^{\bullet}$ and $R^{\bullet}$, where R is any appropriate molecular group, e.g. $C_2H_5$), thio radicals ($RS^{\bullet}$) and nitric oxide ($NO^{\bullet}$).

The peroxyl and hydroxyl radicals were selected from this list for use in the present research. The suitability of another reactive oxygen species, hypochlorous acid, was trialled. Also, the reactive nitrogen species peroxynitrite, although not employed in this research, has been used by other research groups(*3;4*) for similar research. All four species are described below.

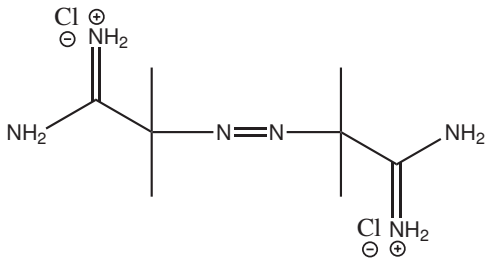
### 7.2.1 Peroxyl Radical

One of the most important radicals in olive oil is the peroxyl radical, as this is formed as an intermediate during lipid oxidation. The main natural pathways of peroxyl radical production are:

(a) Attack by $OH^{\bullet}$ on any organic species. The $OH^{\bullet}$ forms water by abstracting a H atom, leaving a carbon-centred radical. This radical can take on an $O_2$ molecule, thereby forming a peroxyl radical $ROO^{\bullet}$(*2*).

(b) Decomposition of peroxides. The enzyme lipoxygenase produces these peroxides from fatty acids(*5*). They then readily lose a H atom to become peroxyl radicals.

In the present research, an alternative method for production of peroxyl radicals was used: an azo initiator. 2,2'-azobis(2-amidinopropane)hydrochloride (AAPH, figure 7.1) is dissolved in the reaction mixture and maintained at 37°C. This temperature permits the steady decomposition of AAPH, whereby $N_2$ is released and carbon-centred radicals are generated which then add oxygen in the same way as the carbon-centred radicals left after $OH^{\bullet}$ attack.



**AAPH**

**Figure 7.1.** The structure of 2,2'-azobis(2-amidinopropane)hydrochloride, also called AAPH. It is a common 'azo initiator', which means it is an azo compound which readily produces radicals to initiate radical chain reactions.

### 7.2.2 Hydroxyl Radical

Hydroxyl radicals can cause a great deal of damage to any tissue *in vivo*, as they are so reactive they generally react with what ever species is closest to their point of formation. There are several important methods of hydroxyl radical generation:

(a) Fenton chemistry(*6*). This is cleavage of the O-O bond in hydrogen peroxide by $Fe^{2+}$. $OH^{\bullet}$ and $OH^-$ are both produced from this reaction.

(b) Radiation-induced decomposition of appropriate species(*2*). Ultraviolet radiation can be used to homolytically cleave the O-O bond in hydrogen peroxide, and higher energy γ radiation can cleave one of the bonds in water also to form $OH^{\bullet}$.

(c) HOCl, which is formed in the human body in small amounts can react with superoxide ($O_2^{\bullet -}$) to produce $OH^{\bullet}$ along with $Cl^-$ and $O_2$.

For the present research, hydroxyl radicals were formed via a Fenton reaction. Here the C-O bonds in ascorbic acid are cleaved by $Fe^{3+}$. As ascorbic acid

is used for radical generation, the $OH^{\bullet}$ radical species is not ideal for the analysis of ascorbic acid-containing samples.

### 7.2.3 Hypochlorous Acid

Hypochlorous acid (HOCl), while not a radical itself, produces radicals as break-down products. HOCl is formed in humans by the enzyme myeloperoxidase from $H_2O_2$ and $Cl^-$, and is one of several species used during an immune response to destroy pathogens(*2*). HOCl is highly reactive, and as such is not very specific in its reactions.

### 7.2.4 Peroxynitrite

Like hypochlorous acid, peroxynitrite itself is not a radical. Peroxynitrite is the product of reaction between nitric oxide ($NO^{\bullet}$) and superoxide ($O_2^-$). $NO^{\bullet}$ is formed from L-arginine by nitric oxide synthase enzymes(*7*) and also from nitrites by stomach acid(*8*). In humans it is involved with regulation of blood pressure, among other biological functions. Whereas $NO^{\bullet}$ is a necessary radical which is relatively slow to react with non-radicals, peroxynitrite ($ONOO^-$) is a very reactive species. It will react with many different classes of molecules, but only if they are in close proximity to the site of its formation, as it quickly rearranges to $NO_3^-$(*2*).

## 7.3 Radical Scavengers in Lipids

There are many classes of compounds in olive oil which are considered to exhibit radical scavenging capacity. The major contributors to the overall capacity have been shown to be polyphenols, which were described in section 1.2 along with other notable olive oil components. Carotenoids and tocopherols are also present and can scavenge radical species under the appropriate conditions.

The reactive species listed above are not represented equally in lipid media. The peroxyl radical is the most prevalent, as it is an intermediate product of lipid oxidation reactions. Therefore, it is important to differentiate between radical scavenging capacity in biological systems and in lipid systems. Antioxidant compounds can be expected to reside in different local environments and scavenge different radicals with different efficiencies in different media(*9*).

## 7.4 Antioxidant Assays

### 7.4.1 Introduction

Many different assays have been developed to measure the antioxidant capacity of a compound or mixture of compounds. Each has its own range of applicability. For example, if a measure of the antioxidant capacity of a compound in human blood is desired, a lipid-based assay would not be appropriate, as blood is not a lipid. The conditions of the system (in this case blood) to which the results will be applied should be recreated as closely as possible in the assay(*9*). This means that if blood is the system to which the results are applied, the same types of radicals and the same types of vulnerable biological molecules as are found in blood should be present for the compound under investigation to be properly evaluated. Antioxidant assays are commonly used to estimate antioxidant capacity in either bulk lipid systems (for indications of shelf-life) or living tissues (for indications of degree of protection from harmful processes involving radical species). However, no *in vitro* assay has yet been developed which sufficiently recreates a biological system to the extent that it can predict the *in vivo* antioxidant capacity of a sample(*9*).

To analyse a lipid sample, many assays involve the extraction of a selection of compounds from the lipid with a solvent such as methanol or acetone, while some operate in the lipid medium itself. These different assays give different results for the same compounds, and the results should not be taken out of context. The lipid-based assays show antioxidant capacity in lipid systems, while aqueous assays do not. There is only one assay at present which can operate in emulsions – the Selected Ion Flow Tube Mass Spectrometry Total Oxyradical Scavenging Capacity (SIFT-MS-TOSC) assay. An emulsion is important for biological relevance, as lipids are emulsified in the blood(*10*), and it is in this environment that the associated antioxidants scavenge radicals. Results from assays which do not use emulsions should be interpreted with caution.

Antioxidant assays can be separated into two classes: those which operate via a hydrogen atom transfer (HAT) mechanism and those which use a single electron transfer (SET) mechanism(*11*). Most assays based on SET mechanisms involve a probe which is reduced directly by any antioxidants present. Examples of these are the Trolox Equivalent Antioxidant Capacity (TEAC) assay and the Folin-Ciocalteu Total Phenols assay. The Total Oxyradical Scavenging Capacity (TOSC) and SIFT-MS-TOSC assays involve an electron transfer reaction when the hydroxyl radical is used, however the antioxidants do not act directly on the molecular probe. The reactions utilised here follow the same principle as assays employing a hydrogen atom transfer mechanism. Assays based on HAT mechanisms involve the generation of a reactive species. The sample's effectiveness at suppressing the normal reactions of the reactive species is evaluated. Examples of this type of assay

are the ORAC, Total Radical-trapping Antioxidant Parameter (TRAP), TOSC, and SIFT-MS-TOSC assays.

The reactions involved in a HAT-based assay are centred around a radical species, commonly formed from an azo initiator ($R_2N_2$), attacking a substrate (L). Denisov and Khudyakov(*12*) give a description of lipid autoxidation reactions. The relevant reactions from this description are shown below. Lipid autoxidation is a radical chain reaction process, and as such it has three stages: initiation (radical generation), propagation (radical attack of the substrate) and termination (radical quenching).

The initiation reactions involve the loss of nitrogen gas which leads to the generation of two carbon-centred radicals (reaction 7.1). Oxygen adds to form a peroxyl radical (reaction 7.2), and a hydrogen atom is taken from the substrate to produce a substrate radical and a peroxide molecule (reaction 7.3).

$$R_2N_2 \rightarrow 2R\bullet + N_2 \quad (7.1)$$

$$R\bullet + O_2 \rightarrow ROO\bullet \quad (7.2)$$

$$ROO\bullet + LH \rightarrow ROOH + L\bullet \quad (7.3)$$

Fatty acids are the substrates for lipid oxidation, as they are by far the most highly concentrated compounds in lipids. However the assays described in this chapter involve different substrates which exhibit different behaviour. Propagation reactions can occur in oxidising lipids. The fatty acid radicals (L•) propagate the

radical reactions by forming peroxyl radicals themselves (reaction 7.4) and abstracting hydrogen atoms from other substrate molecules (reaction 7.5).

$$L\bullet + O_2 \rightarrow LOO\bullet \quad (7.4)$$

$$LOO\bullet + LH \rightarrow LOOH + L\bullet \quad (7.5)$$

Inhibition occurs when the fatty acid peroxyl radical encounters an antioxidant molecule (AH) and accepts a hydrogen atom (reaction 7.6).

$$LOO\bullet + AH \rightarrow LOOH + A\bullet \quad (7.6)$$

The majority of antioxidant assays (including the SIFT-MS-TOSC assay) are not conducted in a lipid medium and do not allow significant radical chain propagation(*1*). This is because each assay uses its own carefully chosen substrate which either is unreactive (ORAC and TRAP) or breaks down (TOSC and SIFT-MS-TOSC) after radical attack (reaction 7.3). More information on these assays is given later in this chapter. For such assays, reactions 7.4-7.6 may be disregarded. The important inhibition steps in these circumstances are shown in reactions 7.7 and 7.8.

$$R\bullet + AH \rightarrow RH + A\bullet \quad (7.7)$$

$$ROO\bullet + AH \rightarrow ROOH + A\bullet \quad (7.8)$$

The antioxidants donate a hydrogen atom and are left with an unpaired electron. If AH is not a very good antioxidant (A• is equally as reactive as R• and ROO•), reactions 7.7 and 7.8 will make little difference to the protection of the substrate (LH, in this case a lipid). On the other hand, if AH is a good antioxidant (A• is more inert than R• and ROO•), reactions 7.9, 7.10 and 7.11 occur. These reactions can provide significant protection for the substrate.

$$A\bullet + A\bullet \rightarrow \text{non-radical products} \quad (7.9)$$

$$A\bullet + R\bullet \rightarrow \text{non-radical products} \quad (7.10)$$

$$A\bullet + ROO\bullet \rightarrow \text{non-radical products} \quad (7.11)$$

The effectiveness of an antioxidant depends critically on how stable it is after hydrogen atom transfer (reactions 7.7 and 7.8). It must be stable enough to react only with other radical species (reactions 7.9, 7.10 and 7.11) if it is to offer protection to the substrate.

### 7.4.2 Folin-Ciocalteu Total Phenols Assay

As described by Huang, Ou and Prior(*1*), this assay measures the reducing capacity of a sample, suggesting that it does not provide a selective measure of antioxidants alone, but will measure any constituent which will accept electrons. Although the method seems quite general on the surface, it appears remarkably selective for compounds containing phenolic hydroxyl groups. The Folin-Ciocalteu assay uses a reagent appropriately named the Folin-Ciocalteu reagent, which is a

chemical cocktail formed from sodium tungstate, sodium molybdate, phosphoric acid and lithium sulfate(*13*), as a molecular probe(*1*).

The exact structure of the reactive species in the Folin-Ciocalteu reagent solution is not known, but it is believed that the colour change from a yellow to a deep blue colour during reaction is caused by the reduction of molybdenum from Mo(VI) to Mo(V), forming $[PMoW_{11}O_{40}]^{4-}$(*1*). This seems to be quantitative, depending on the number of phenolic hydroxyl groups present in the sample. According to Roginsky and Lissi(*14*), phenolics are excited in basic solution, forming $O_2^-$ which is the reacting species. It is not clear what this assertion is based on, as no other articles have been found which include it, and no study is cited which demonstrates it. The exact mechanism for the electronic exchange is unknown, yet the Folin-Ciocalteu assay provides the most widely used and trusted measurement of antioxidant capacity. This of course assumes that phenolic antioxidants are the only important antioxidants in the sample under analysis. New methods are often compared to the Folin-Ciocalteu total phenols assay as evidence of their suitability.

However, the total phenols assay is not specific – the Folin-Ciocalteu reagent is also reduced by vitamin C and the Cu(I) ion(*1*).

### 7.4.3 Total Oxyradical Scavenging Capacity

The TOSC assay was first reported by Winston et. al.(*15*), who used the well-known break down of α-keto-γ-methylthiobutanoic acid (KMBA, the substrate in this assay) by radical species to produce ethene(*16-18*). Winston et. al. employed the reaction as a measure of radical scavenging capacity by antioxidants in a

competitive assay – scavenging more radicals leaves fewer to react with KMBA, causing a lower rate of ethene production. The radical species used to produce peroxyl radicals for the original TOSC assay was 2,2'-azobis-amidinopropane (ABAP), and ethene production was measured by Gas Chromatography with a Flame Ionisation Detector (GC-FID)(*15*). The TOSC assay was later extended to include the use of three radical species: peroxyl radicals, hydroxyl radicals and peroxynitrite(*3*). The rest of the details are very similar to those of the SIFT-MS-TOSC assay, which will now be described.

### 7.4.4 SIFT-MS-TOSC

The TOSC assay was developed for SIFT-MS analysis of aqueous systems from the GC-based method by Winston et. al., and recently has been adapted further by Senthilmohan & McEwan(*19*) to measure oil-in-water emulsions. Peroxyl radicals are produced by thermal decomposition of AAPH at 37°C. Reaction of these radicals with KMBA releases ethene (figure 7.2)(*15*). The other two oxidising species from the TOSC assay(*3*), $OH^{\bullet}$ and peroxynitrite, are produced in the same way in the SIFT-MS-TOSC assay as in the original TOSC assay, i.e. via a Fenton reaction and decomposition of 3-morpholinosydnonimine N-ethylcarbamide(*3*). The peroxyl and hydroxyl radicals only were used in the present research.

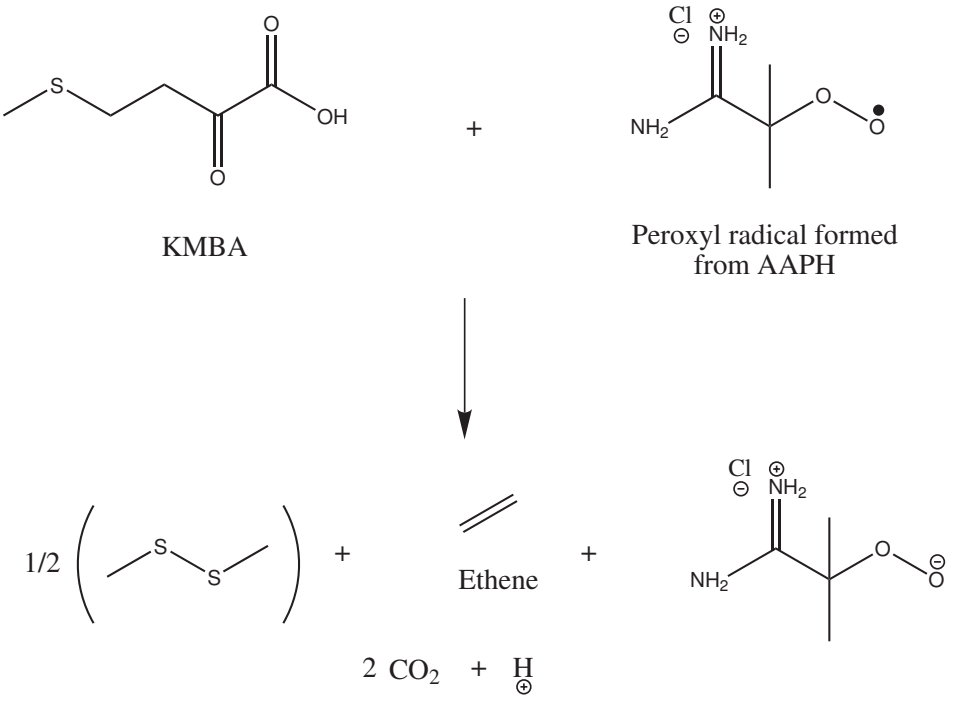



**Figure 7.2.** The reaction between KMBA and a peroxyl radical (ROO•) from AAPH forms ethene along with several other products.

The reaction of hydroxyl radicals with KMBA is thought to proceed via single electron transfer from the sulfide group(*17*), however no reaction mechanism has been proposed for the reaction of peroxyl radicals with KMBA. The measured products are the same for both reactions, therefore the mechanisms are expected to be similar. It should be noted however that the hydroxyl radical reacts via electron abstraction, while the peroxyl radical generally reacts via hydrogen atom transfer. Whatever the mechanism, radical attack causes the KMBA molecule to cleave, giving two molecules of carbon dioxide, one of ethene and a methylthio radical (two of which may combine to produce dimethyl disulfide)(*17*). The production of acetone (the concentration of which is also observed to increase during the reaction)

was not described by Winston et. al.(*15*), and is thought to be due solely to AAPH break-down products. The ethene concentration in the reaction vial headspace is monitored by SIFT-MS from its reaction with $O_2^+$ (reaction 7.12).

$$k = 1.0 \times 10^{-9}\ cm^3s^{-1}$$

$$O_2^+ + C_2H_4 \rightarrow C_2H_4^+ + O_2 \qquad (7.12)$$

The important reactions in the TOSC assay occur in the aqueous phase, as both the KMBA substrate and the radicals derived from AAPH are highly soluble in water. To include olive oil, an oil-in-water emulsion is used. The hydrophilic components of olive oil dissolve into the aqueous phase from the emulsion droplets, while the hydrophobic components remain. These hydrophobic compounds are still able to protect the substrate, as they can scavenge radicals at the oil-water interface(*20*).

The method details are described in section 8.1. In brief, several different concentrations of oil are used in different bottles, giving a different inhibition value for each concentration. Using these inhibition values, it is possible to obtain the oil concentration which will cause 50 % inhibition of ethene production compared to the blank solution without olive oil. This value is called the Inhibition Concentration at 50 % ($IC_{50}$). Due to the variability in composition of olive oil, a reliable measure of its concentration in mol $L^{-1}$ is difficult to obtain. In the present research, the volume necessary to produce 50 % inhibition of ethene was found as it

is much easier to calculate and is considered to be proportional to the concentration to within experimental error.

The concentrations of AAPH and KMBA chosen are appropriate to give a good head space ethene concentration over the hour that the reaction is allowed to proceed. The range of concentrations of emulsified oil in the system is currently based on an estimate. A series of oil volumes are chosen which are then used to estimate the volume at which 50 % inhibition is achieved. This volume is termed the $VI_{50}$ value. If the $VI_{50}$ value is outside the chosen concentration range of the oils, the assay must be repeated, as the line of best fit is not always linear and extrapolation may be unreliable.

The SIFT-MS-TOSC assay has potential as a model for *in vivo* radical scavenging, where oil is emulsified and reactions occur in the aqueous phase, so the relative contributions of different antioxidant species in the SIFT-MS-TOSC assay are as close as currently possible to those *in vivo*. There are also different radical species which may be used, approximating this aspect of biological systems. However, the SIFT-MS-TOSC assay, although much better suited to rapid analysis than the TOSC assay, does have some difficulty with high-throughput analysis. Also, while one of the best substrates currently available is used (KMBA, which is similar to the naturally occurring amino acid methionine, the suspected source of ethene in plants(*17*)), the biological relevance of an assay which measures the ability to protect only KMBA is limited.

### 7.4.5 Concluding Remarks

Many different assays exist for assessing antioxidant capacity, but only a selection of the most widely used assays have been mentioned above. In 2004 the first of a series of annual meetings on antioxidant methods was held with the purpose of choosing an analytical method or methods to best measure antioxidant capacity in foods and dietary supplements(*21*). A number of good review articles have also been written on the subject(*9;11*). As no single assay adequately recreates the *in vivo* situation, Frankel and Meyer(*9*) suggest the use of several standardised assays, as each assay has its own potential for interference and bias towards measuring certain types of antioxidants. Analysing the same sample with several different assays gives a better appreciation of the true *in vivo* antioxidant capacity. The aim of this research was to develop a SIFT-MS-TOSC assay that could be used for a wide range of non-aqueous samples and to evaluate its use in the analysis of olive oil antioxidants.

## 7.5 References


(1) Huang, D.; Ou, B.; Prior, R. The Chemistry Behind Antioxidant Capacity Assays. *J. Agric. Food Chem.* **2005**, *53*, 1841-1856.

(2) Halliwell, B.; Gutteridge, J. *Free Radicals in Biology and Medicine*, 3rd ed.; Oxford University Press: Chennai, India, 1999.

(3) Regoli, F.; Winston, G. W. Quantification of Total Oxidant Scavenging Capacity of Antioxidants for Peroxynitrite, Peroxyl Radicals, and Hydroxyl Radicals. *Toxicol. Appl. Pharm.* **1999**, *156*, 96-105.

(4) Lichtenthäler, R.; Marx, F.; Kind, O. M. Determination of Antioxidative Capacities Using an Enhanced Total Oxidant Scavenging Capacity (TOSC) Assay. *Eur. Food Res. Technol.* **2003**, *216*, 166-173.

(5) Belitz, H.-D.; Grosch, W. *Food Chemistry*, 2nd English ed.; Springer-Verlag: Berlin, Germany, 1999.

(6) Uri, N. Inorganic Free Radicals in Solution. *Chem. Rev.* **1951**, *50*, 375-454.

(7) Knowles, R. G.; Moncada, S. Nitric Oxide Synthases in Mammals. *Biochem. J.* **1994**, *298*, 249-258.

(8) Halliwell, B. What Nitrates Tyrosine? Is Nitrotyrosine Specific as a Biomarker of Peroxynitrite Formation *In Vivo*? *FEBS Letters* **1997**, *411*, 157-160.

(9) Frankel, E. N.; Meyer, A. S. The Problems of Using One-Dimensional Methods to Evaluate Multifunctional Food and Biological Antioxidants. *J. Sci. Food Agric.* **2000**, *80*, 1925-1941.

(10) Mathews, C. K.; van Holde, K. *Biochemistry*; The Benjamin/Cummings Publishing Company: Redwood City, CA, USA, 1990.

(11) Prior, R.; Wu, X.; Schaich, K. Standardized Methods for the Determination of Antioxidant Capacity and Phenolics in Foods and Dietary Supplements. *J. Agric. Food Chem.* **2005**, *53*, 4290-4302.

(12) Denisov, E. T.; Khudyakov, I. V. Mechanisms of Action and Reactivities of the Free Radicals of Inhibitors. *Chem. Rev.* **1987**, *87*, 1313-1357.

(13) Folin, O.; Ciocalteu, V. On Tyrosine and Tryptophane Determinations in Proteins. *J. Biol. Chem.* **1927**, *73*, 627-650.

(14) Roginsky, V.; Lissi, E. A. Review of Methods to Determine Chain-Breaking Antioxidant Activity in Food. *Food Chem.* **2005**, *92*, 235-254.

(15) Winston, G.; Regoli, F.; Dugas, A. J.; Fong, J.; Blanchard, K. A Rapid Gas Chromatographic Assay for Determining Oxyradical Scavenging Capacity of Antioxidants and Biological Fluids. *Free Rad. Biol. Med.* **1998**, *24*, 480-493.

(16) Lieberman, M.; Kunishi, A.; Mapson, L.; Wardale, D. Ethylene Production from Methionine. *Biochem. J.* **1965**, *97*, 449-459.

(17) Yang, S. Further Studies on Ethylene Formation from α-Keto-γ-Methylthiobutyric Acid or β-Methylthiopropionaldehyde by Peroxidase in the Presence of Sulfite and Oxygen. *J. Biol. Chem.* **1969**, *244*, 4360-4365.

(18) Beauchamp, C.; Fridovich, I. A Mechanism for the Production of Ethylene from Methional. *J. Biol. Chem.* **1970**, *245*, 4641-4646.

(19) Senthilmohan, S.; McEwan, M. A Method of Assaying the Antioxidant Activity of Pure Compounds, Extracts and Biological Fluids. New Zealand, 2005.

(20) Frankel, E.; Huang, S.-W.; Kanner, J.; German, J. Interfacial Phenomena in the Evaluation of Antioxidants: Bulk Oils vs Emulsions. *J. Agric. Food Chem.* **1994**, *42*.

(21) Conference Targets Uniform Antioxidant Measurements. http://www.chemistry.org/portal/a/c/s/1/feature_acs.html?id=c373e9fcee851a518f6a4fd8fe800100 (accessed 15 Aug, 2005)

# Chapter 8

## TOSC Assay Development

### 8.1 SIFT-MS-TOSC Assay Method

#### 8.1.1 Introduction

The SIFT-MS-TOSC assay requires the preparation of a mixture at the correct pH and temperature, containing each component at the correct relative concentration. The assay provides a measure of an oil's ability to scavenge radical species. To perform the SIFT-MS-TOSC assay using peroxyl radicals buffer, DTPA, surfactant, KMBA and AAPH, as well as the olive oil to be evaluated are included in the mixture. For the assay using hydroxyl radicals buffer, ascorbic acid, surfactant, KMBA and AAPH are included with the olive oil. All of these compounds and the preparation of their respective mixtures will now be described.

#### 8.1.2 Buffer solution

The SIFT-MS-TOSC assay is performed in a $H_2PO_4^-/HPO_4^{2-}$ buffer system, with pH = 7.4. The potassium dihydrogen phosphate solution contains $13.61 \pm 0.01$ g $L^{-1}$ $KH_2PO_4$ (AnalaR grade, BDH Laboratory Supplies, Poole, England), while the sodium monohydrogen phosphate solution contains $35.82 \pm 0.01$ g $L^{-1}$ $Na_2HPO_4.12H_2O$ (AnalaR grade, BDH Laboratory Supplies, Poole, England). The buffer is then made up with $19.6 \pm 0.4$ % by volume of the potassium dihydrogen phosphate solution and $80.4 \pm 0.4$ % by volume of the sodium monohydrogen phosphate solution.

Dissolved iron may affect the assay results, therefore a ligand was included to bind the iron and prevent its interference in this assay. The ligand used was diethylenetriaminepentaacetic acid (DTPA, ≥ 99 %, Sigma-Aldrich Co., St. Louis, MO, USA). DTPA was added to the buffer solution before inclusion in the assay mixture. The volume of buffer added to the assay mixture varied depending on the volume of oil analysed, yet the concentration of DTPA required was always the same, therefore the buffer was divided into two different solutions. The same volume of the solution containing DTPA was always added, while the other solution was solely phosphate buffer and its volume was altered to keep the overall assay mixture at a constant volume. The buffer containing DTPA was prepared by adding $78.8 \pm 0.4$ mg $L^{-1}$ of DTPA to the phosphate buffer solution.

### 8.1.3 Surfactant solution

A surfactant was used to emulsify the olive oil in the assay mixture. This surfactant solution was prepared by adding $20.00 \pm 0.05$ g $L^{-1}$ of Pluronic P 104 surfactant (BASF Corporation, Mount Olive, NJ, USA) to the buffer solution. This was left for at least four hours (usually overnight) to allow proper dissolution of the surfactant.

### 8.1.4 Emulsion

To prepare the emulsion, $2.00 \pm 0.03$ mL of oil was maintained in a measuring cylinder at $55 \pm 2$ °C in a water bath for several minutes, while $18.0 \pm 0.5$ mL of surfactant solution was maintained at $45 \pm 2$ °C in a water bath for the same length of time. The surfactant solution was added to the oil in the measuring cylinder, and immediately mixed with a D-500 homogeniser

(Wiggenhauser, Berlin, Germany) at 1 x $10^4$ rpm for 10.00 ± 0.05 min while being maintained at 45 ± 2 °C in a water bath. Immediately after mixing, the emulsion was cooled to room temperature in a water bath.

### 8.1.5 Substrate for Radical Reaction

The substrate in the SIFT-MS-TOSC assay produces ethene upon reaction with radicals, thereby allowing its rate of reaction to be measured. The substrate used for radical reaction was α-keto-γ-methylthiobutanoic acid (KMBA). KMBA solution (2.30 x $10^{-2}$ mol $L^{-1}$) was prepared by dissolving 3.45 ± 0.05 mg KMBA (sodium salt, Sigma-Aldrich Co., St. Louis, MO, USA) in 1.00 ± 0.01 mL of distilled water.

### 8.1.6 Peroxyl radical generator

To produce radicals for reaction with the substrate, 2,2'-Azobis(2-amidinopropane)hydrochloride (AAPH) solution (2.00 x $10^{-1}$ mol $L^{-1}$) was prepared by dissolving 542.5 ± 0.5 mg AAPH (97%, Aldrich Chemical Company, Inc., Milwaukee, WI, USA) in 10.00 ± 0.02 mL buffer solution. As AAPH is unstable in aqueous solution, this was prepared last, just before the start of the experiment.

### 8.1.7 Peroxyl assay mixture

The total volume of solution in each 1 L bottle (Schott Glass, Mainz, Germany) was 10 mL, made up of 5.00 ± 0.02 mL buffer containing DTPA, 0.100 ± 0.001 mL KMBA solution, 1.00 ± 0.01 mL AAPH solution, the desired amount of oil-in-water emulsion and the rest made up with the buffer solution. The

AAPH was always added last, to coincide with the beginning of the assay. Typical concentrations of the reagents in the assay are shown in table 8.1.

**Table 8.1.** Typical concentrations of components of the SIFT-MS-TOSC assay mixture in the peroxyl and hydroxyl radical assays. Due to the variable composition of olive oil and the variable composition of the surfactant, these are expressed by volume and mass respectively. The hydroxyl radical assay components are described below.

| Peroxyl radical assay | | | Hydroxyl radical assay | | |
|---|---|---|---|---|---|
| Olive oil | 50 | µL | Olive oil | 50 | µL |
| Surfactant | 0.9 | g $L^{-1}$ | Surfactant | 0.9 | g $L^{-1}$ |
| DTPA | 1.00E-04 | mol $L^{-1}$ | Ascorbic acid | 1.80E-04 | mol $L^{-1}$ |
| KMBA | 2.30E-04 | mol $L^{-1}$ | KMBA | 2.30E-04 | mol $L^{-1}$ |
| AAPH | 2.00E-02 | mol $L^{-1}$ | $FeCl_3$ | 1.80E-06 | mol $L^{-1}$ |
| | | | EDTA | 4.60E-06 | mol $L^{-1}$ |

### 8.1.8 Hydroxyl radical generator and assay mixture

For the hydroxyl radical assay, ferric ions are used to cleave hydroxyl radicals from ascorbic acid via the Fenton reaction (see section 7.2.2). The rate of •OH release is controlled by the addition of EDTA, which limits the concentration of free ferric ion, and therefore restricts ascorbic acid from binding to the ferric ions.

100 ± 1 µL of 8.93 ± 0.05 mol $L^{-1}$ $FeCl_3$ was made up to 10.00 ± 0.04 mL with distilled water in a volumetric flask. 40 ± 1 µL of this solution was then added to a 100 mL volumetric flask. 10.7 ± 0.1 mg EDTA was made up to 5.00 ± 0.02 mL with distilled water in a volumetric flask. 1.25 ± 0.02 mL of this solution was added

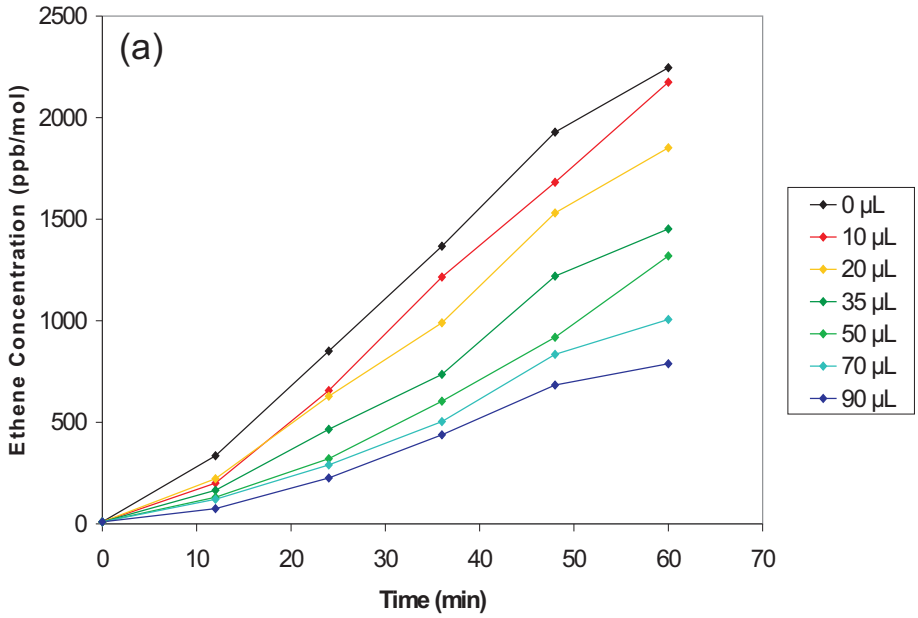
to the same 100 mL volumetric flask as the $FeCl_3$ solution, and made up to 100.0 ± 0.8 mL with the phosphate buffer.

The hydroxyl radical assay does not require the addition of DTPA to the reaction mixture. It does, however, require ascorbic acid to be decomposed by the ferric ions to produce the hydroxyl radicals. The ascorbic acid is added to the phosphate buffer in the same way as the DTPA in the peroxyl radical assay: the buffer solution is split in two, with one solution containing ascorbic acid and the other without. The same volume of the buffer containing ascorbic acid was always added to the assay mixture, while the volume of the lone buffer was adjusted so that the total assay mixture was constant across all samples. 7.9 ± 0.1 mg of ascorbic acid was dissolved in 100.0 ± 0.8 mL phosphate buffer and 4.00 ± 0.03 mL of this solution was added to each assay bottle. As in the peroxyl assay, the total mixture volume was 10 mL. The balance after accounting for the buffer containing ascorbic acid, emulsion, KMBA and $FeCl_3$/EDTA was made up with the phosphate buffer which did not contain ascorbic acid. 500.0 ± 0.6 μL of the $FeCl_3$/EDTA solution was added last to each 1 L bottle to commence the assay. Typical concentrations of the reagents in the assay are shown in table 8.1.

### 8.1.9 Analysis technique

Seven bottles were used per experiment: one control and six with increasing concentrations of oil emulsion. All bottles were kept at 37 ± 0.2 °C in a water bath throughout the experiment. Each bottle was analysed for its head space ethene concentration every twelve minutes for a total of 5 times, making a total monitoring time of one hour. The head space volume extracted during each measurement

(approximately 20-30 mL) was replaced with air immediately after. As this was only 2-3 % of the total bottle head space and the control was undergoing the same procedure as the samples, dilution of the head space in this way had little effect on the final assay results. The concentration of ethene in each bottle for every 12 minute segment (figure 8.1(a)) was calculated as an area (time multiplied by the average of concentrations at the beginning and end of the interval). The total area produced for each bottle was subtracted from the control to give a percent inhibition. The percent inhibition displayed in each bottle was plotted against the volume of oil added to the corresponding bottle (figure 8.1(b)) and the resultant function used to obtain the volume of oil necessary to cause a 50 % drop in (or inhibition of) head space ethene concentration in a sample bottle relative to the control. This value is known as the 50 % inhibition volume, or $VI_{50}$.



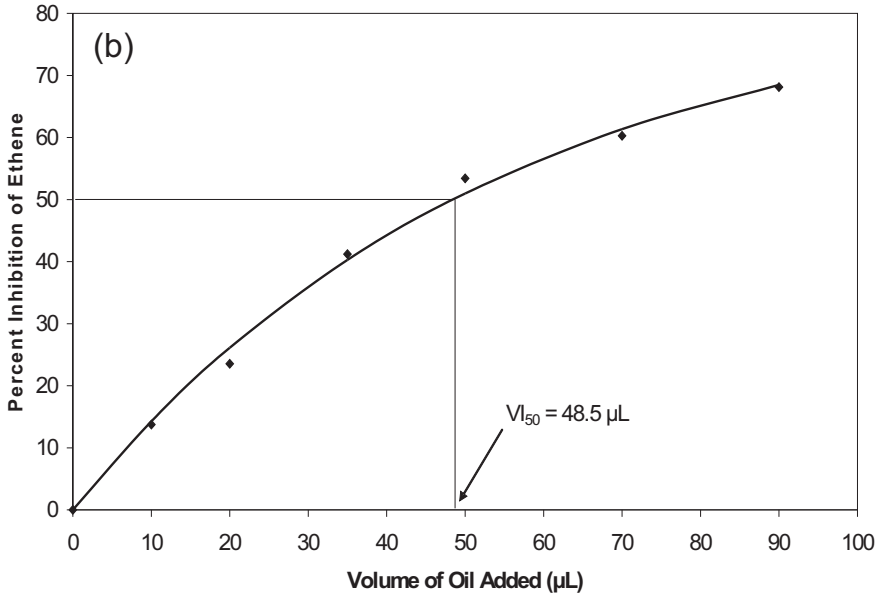



**Figure 8.1.** Results from the SIFT-MS-TOSC assay for an olive oil sample. Raw results **(a)** consist of ethene concentration data over the course of an hour for six bottles containing different added volumes of a single olive oil. To determine the $VI_{50}$, the area under the curve for each bottle is calculated as a percent inhibition of ethene production (1 – sample area / area under control curve). The data are plotted as percent inhibition of ethene against volume of oil added **(b)** and a function is fitted to approximate the data. The added volume necessary to achieve 50 % inhibition (the $VI_{50}$) is then found.

## 8.2 SIFT-MS-TOSC Assay Development

### 8.2.1 Incorporating HOCl into the Assay

Experiments involving HOCl were conducted in the hope of adding it to the list of possible reactive species that can be included in the SIFT-MS-TOSC assay. A range of different ratios of HOCl to KMBA concentrations were used. Two

different reactions were expected: loss of Cl• followed by reaction of HO• with KMBA to produce ethene(*1*), and oxidation of the thioether group in KMBA to a sulfoxide group(*2*). The latter reaction would not produce ethene, so it was hoped this would be a minor reaction channel.

Very low, yet significant concentrations of ethene were detected during all five attempts, although with poor repeatability (figure 8.2). These results suggest that the dominant reaction pathway does not produce ethene. However, the fact that low concentrations of ethene were observed indicates that the desired reaction is occurring, only too slowly to be of use in the SIFT-MS-TOSC assay.

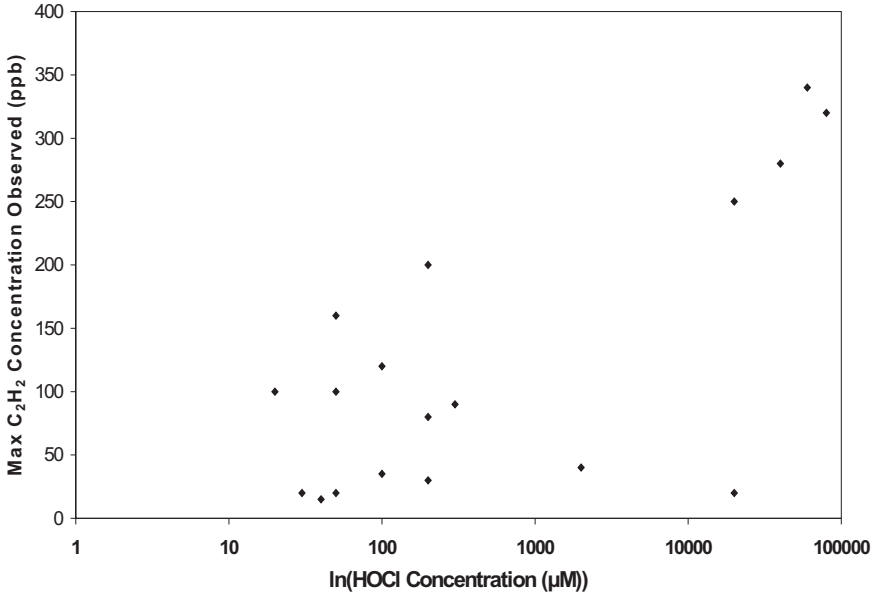



**Figure 8.2.** Results from the HOCl SIFT-MS-TOSC assay trial. The ethene concentrations observed are neither high enough nor reproducible enough to provide accurate and precise evaluation of antioxidant capacity.

A common method of preventing unwanted side-reactions is to add a protecting group at the reactive site. In this case it would be added to the sulfur of the KMBA molecule. However, as the desired reaction (ethene production) occurs at the same location as the undesired reaction (the sulfur) and is also an oxidation step, protecting the sulfur would halt all reaction with KMBA. In order to properly measure HOCl, the use of a substrate other than KMBA is required. HOCl reacts rapidly with several different functional groups(*2*). An appropriate substrate for this assay should react selectively and rapidly with HOCl to release characteristic volatile products. This work is beyond the scope of the present research and was not pursued.

### 8.2.2 Preparing the SIFT-MS-TOSC Assay for Routine Use

The SIFT-MS-TOSC assay as described previously displays several shortcomings:

- only one oil sample is able to be analysed using one radical species per assay
- substantial manual effort is needed to prepare and perform the assay.

To simplify the assay procedure it was decided to generate a reference function which would relate a chosen added volume of an olive oil to its $VI_{50}$ value. To achieve this, 16 assays of different olive oil emulsions were performed using the peroxyl radical, spanning a wide range of inhibition values. Six different oil volumes were used in each assay, enabling the degree of inhibition relative to the control to be calculated for each volume. The oil volume used in each bottle in each assay was divided by the $VI_{50}$ value derived for that particular oil. All points for each assay now describe a function which reaches the $VI_{50}$ value at an arbitrary

volume of 1. This alteration allows the direct comparison of all points for all the peroxyl assays with each other. The resultant chart is shown in figure 8.3. To obtain the absolute volume of olive oil corresponding to any point from the chart, only the $VI_{50}$ value for that particular oil is needed.

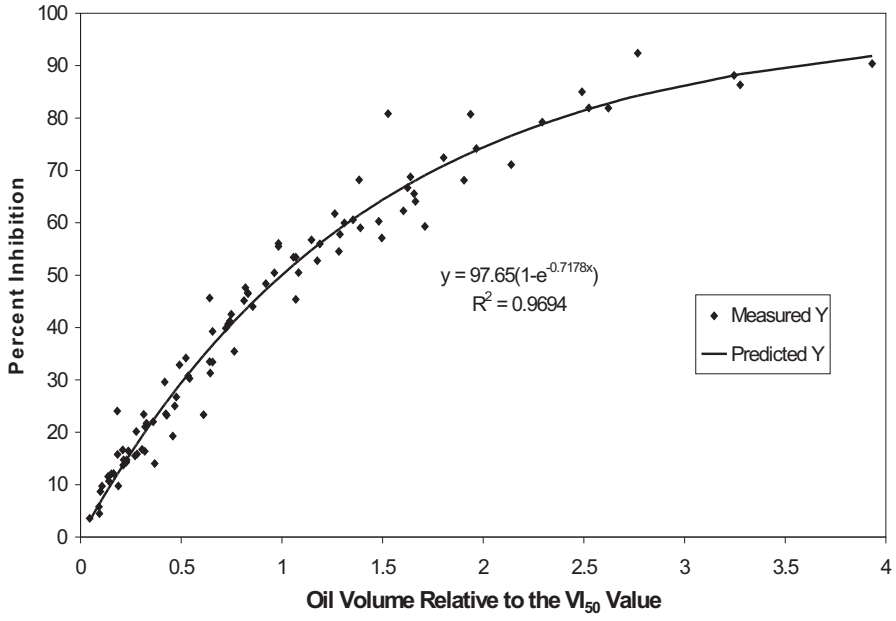



**Figure 8.3**. Inhibition scores for 16 peroxyl radical assays of emulsified olive oil. Each assay was linearly adjusted so as to exhibit 50 % inhibition at the arbitrary value of 1 and hence to be directly comparable with the other assays conducted. The inverse exponential function fitted to these data is henceforth called the 'peroxyl reference function.'

Good agreement is shown between the assays here, suggesting that a reliable estimate of the $VI_{50}$ value may be obtained from the percent inhibition observed at other known oil volumes – given that the inhibition observed at this volume is not too distant from 50 %. The distribution is close to linear up until

approximately 50 % inhibition, with the entire data set approximated well by an inverted exponential decay function. A linear calibration function would of course be preferable to an exponential, however Lichtenthäler et. al.(*3*) and MacLean et. al.(*4*) have also reported non-linear responses with concentration from the TOSC assay, supporting the non-linearity of the TOSC assay response. Lichtenthäler did not state the function type used to approximate the observed curve, while MacLean fitted a second degree polynomial.

The inverted exponential function used to model the effect of increasing sample volume on the inhibition of ethene production observed in the SIFT-MS-TOSC assay was of the form (equation 8.1):

$$y = A\left(1 - e^{-Bx}\right) \qquad (8.1)$$

The use of only one oil volume for the determination of $VI_{50}$ for each oil is much more efficient in terms of both time and effort than using six different concentrations of each oil emulsion. As the SIFT-MS-TOSC assay was carried out in this study using a single water bath which held seven bottles, a control and triplicate measurement for two samples are possible. The number of samples able to be analysed per assay is therefore doubled by the use of the peroxyl reference function.

The success of the peroxyl reference function for the peroxyl radical assay led to the same investigation using the hydroxyl radical. Ten assays were performed, spanning a large range of inhibition values. The resultant chart when

these volume data were divided by their corresponding $VI_{50}$ values is shown in figure 8.4.

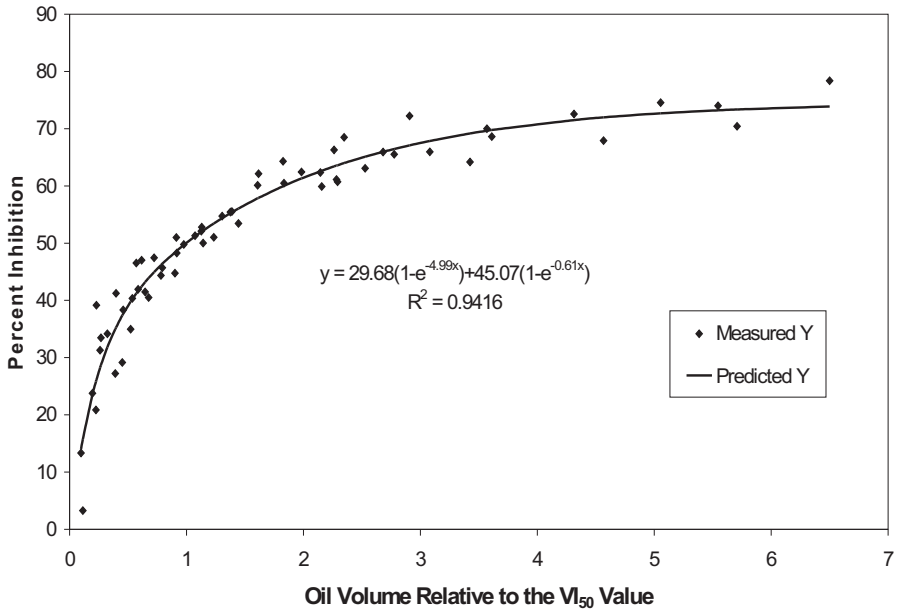



**Figure 8.4**. The same treatment of the hydroxyl radical SIFT-MS-TOSC assay as is seen for the peroxyl radical in figure 8.3 is shown here, involving ten assays. Note that this function was also constructed to reach 50% inhibition at the arbitrary volume of 1.

The hydroxyl radical assay results show a good fit but a slightly lower $R^2$ value than for the peroxyl radical assay. However, a different function is required to obtain a satisfactory fit for the data. The double exponential shown here may be replaced by a logarithmic function, although the double exponential provides a better fit and is of a similar type to the peroxyl reference function.

Consideration of the reactions undergone during the SIFT-MS-TOSC assay gives some insight into why the inhibition observed at different antioxidant

concentrations does not describe a linear function. Due to the nature of radical production used in the SIFT-MS-TOSC assay, the assay reactions cannot be modelled by rate equations. The steady rate at which radicals are produced is the rate-determining step, with all other reactions occurring almost instantaneously. Therefore to model the reaction, we need only consider the probability of the reaction of radicals or reactive agents with all radical scavengers that are present. This will show what modifications in ethene production we can expect with increasing antioxidant concentration. Each radical produced reacts with one of two species. The first is KMBA and it is this reaction that is measured via the amount of ethene generated. The second is an antioxidant. The antioxidant reduces the amount of ethene produced and this is apparent from a comparison of a sample (antioxidant present) and the control bottle (no antioxidant). In this simple scheme, the probability of ethene being produced ($P_{ethene}$) by a radical is given by equation 8.2.

$$P_{ethene} = \frac{k_{KMBA}[KMBA]}{k_{AO}[AO] + k_{KMBA}[KMBA]} \qquad (8.2)$$

where [AO] is the total antioxidant concentration and $k_{KMBA}$ and $k_{AO}$ are reaction constants reflecting the efficiency of reaction with the radical. $P_{ethene}$ is simply the probability of a radical reacting with KMBA divided by the probability of reacting with all radical scavenging species present. $P_{ethene}$ is proportional to the head space ethene concentration in the assay bottles. Inhibition is defined in the

SIFT-MS-TOSC assay as the fraction of radicals that are prevented from reacting with KMBA. One minus $P_{ethene}$ gives the expected inhibition, which can be converted to percent inhibition. Assigning arbitrary values to the necessary variables ($k_{KMBA} = k_{AO} = 1$, [KMBA] = 1) and plotting the inhibition against the antioxidant concentration (figure 8.5) generates a curve similar to those observed for the real samples. The inhibition function generated from equation 8.2 ($y = 100(1 - \frac{1}{x+1})$ ) does not fit the experimental data as closely as the selected exponential functions (even when $k_{KMBA}$ and $k_{AO}$ values are changed), so there is clearly more occurring during the assay than accounted for in this explanation. However, this simple scheme does show that the non-linear nature of the inhibition response curve is to be expected.

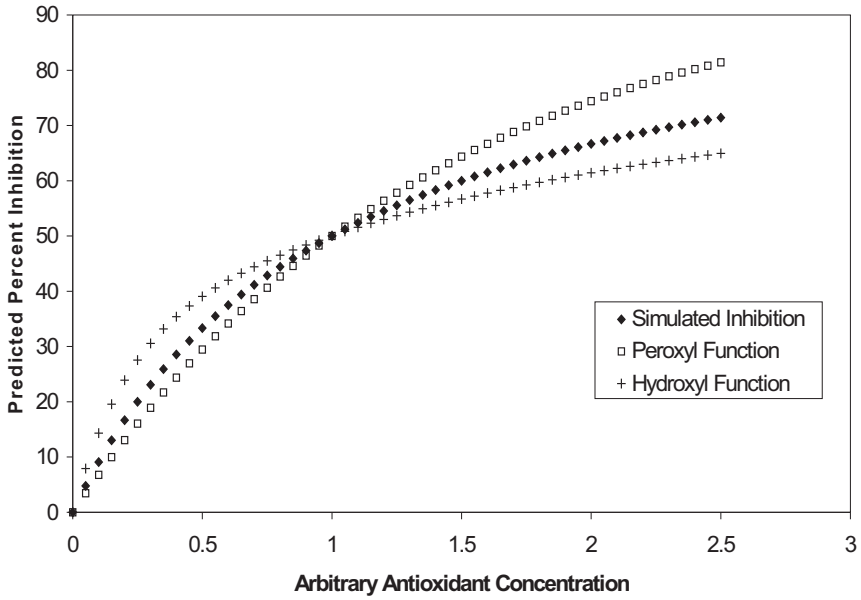



**Figure 8.5.** Simulated inhibition in the SIFT-MS-TOSC assay for a model system compared with functions of best fit derived for the peroxyl and hydroxyl radical SIFT-MS-TOSC assays. The simulated response fits neither of the observed responses, yet lies in between the two.

The degree of work associated with the assay has not been substantially reduced by the analysis of a single mixture composition except that it does provide the option to analyse twice the number of samples with little increase in effort. However, now that identical volumes of all assay components are used in all but the control bottle, the need to decide emulsion volumes and calculate remaining volumes for other solutions has been removed. This change does offer the option of automation for the SIFT-MS-TOSC assay. A simple multiple valve injection system with appropriate software control would minimise manual effort during the assay itself. However, this is outside the scope of the present research. The best

way to minimise time and effort during assay preparation is to prepare enough of the necessary solutions for several assays at one time and either refrigerate or freeze the remainder for future assays. The buffer and its stock solutions may be refrigerated for several months with no noticeable change in their effectiveness during the assay. The surfactant solution and the buffer with DTPA included may be treated in the same way. AAPH and KMBA, however, are not stable enough to refrigerate. Other researchers have frozen solidified solutions of these compounds and successfully employed them in the standard TOSC assay(*3*). The preparation of AAPH and KMBA solutions is not very time-consuming, however, and these reagents are better stored refrigerated in their powder forms. Therefore, the AAPH and KMBA solutions were prepared fresh before each SIFT-MS-TOSC assay. Likewise, the buffer with ascorbic acid employed in the hydroxyl radical assay is not stable enough to be kept refrigerated for any length of time – it was also made fresh before each hydroxyl radical assay as its preparation is not very time-consuming. The $FeCl_3$/EDTA solution is time-consuming to prepare and a large proportion is wasted if it is not stored, so several vials were frozen for future assays after preparation.

The SIFT-MS-TOSC assay is now in an acceptable optimised form for the determination of antioxidant capacity against both peroxyl and hydroxyl radicals. Investigations similar to those described above may be employed to allow the inclusion of other radical species in the optimised assay.

## 8.3 References

(1) Beauchamp, C.; Fridovich, I. A Mechanism for the Production of Ethylene from Methional. *J. Biol. Chem.* **1970**, *245*, 4641-4646.
(2) Winterbourn, C. C.; Kettle, A. J. Biomarkers of Myeloperoxidase-Derived Hypochlorous Acid. *Free Rad. Biol. Med.* **2000**, *29*, 403-409.
(3) Lichtenthäler, R.; Marx, F.; Kind, O. M. Determination of antioxidative capacities using an enhanced total oxidant scavenging capacity (TOSC) assay. *Eur. Food Res. Technol.* **2003**, *216*, 166-173.
(4) MacLean, D.; Murr, D.; DeEll, J. A modified total oxyradical scavenging capacity assay for antioxidants in plant tissues. *Postharvest Biol. Technol.* **2003**, *29*, 183-194.

# Chapter 9

## SIFT-MS-TOSC Assay Characterisation

### 9.1 Introduction

In this chapter some results of the SIFT-MS-TOSC assay for antioxidant capacity on olive oils are presented. There are a number of different antioxidant assays, with each displaying its own bias towards certain classes of compounds or functional groups and giving a higher response to some than to others. The factors governing these responses are not always relevant to the intended applications. As natural sources of antioxidants such as olive oil contain a mixture of different types of antioxidants, different assays may give different results for the same sample, depending on what factors govern the response of each assay. Determining what is actually being measured is important if an assay is to be widely used. The present chapter describes some attempts to discover what aspects of olive oil antioxidants are measured by the SIFT-MS-TOSC assay.

### 9.2 Method

#### 9.2.1 Folin-Ciocalteu Total Phenols Assay

As mentioned in section 7.4.2, the Folin-Ciocalteu assay is particularly selective toward phenolic compounds. The Folin-Ciocalteu assay must be conducted in aqueous solution, therefore the phenolic compounds in olive oil samples were extracted before analysis using the method of Tovar et. al.(*1*) as described next. The assay procedure used was that reported by Scalbert et. al.(*2*). $25 \pm 0.5$ mL olive oil was shaken vigorously with $10 \pm 0.5$ mL of 80:20 v/v methanol/water. The mixture was centrifuged for 10 min at 1900 rpm, and the process repeated using the same olive oil with another 10 mL

methanol/water extraction. The methanol/water extractions were combined to give 20 mL. No further purification steps were considered necessary, as the extract at this point in the method did not produce any interference in the colorimetric assay. 100 ± 4 μL of the extract was added to a 14 mL glass vial. 400 ± 4 μL of 80:20 v/v methanol/water was added to dilute the phenolic extract in the ratio 1:4. 2.50 ± 0.02 mL of the Folin-Ciocalteu reagent (Sigma-Aldrich Co., St. Louis, MO, USA) diluted 1:9 with water was also added. The vial stood for 8.0 ± 0.1 min with occasional shaking, after which 2.00 ± 0.02 mL of 75.0 ± 0.2 g $L^{-1}$ sodium carbonate solution was added. The reaction mixture was transferred to a water bath at 50 ± 2 °C for 5.0 ± 0.1 min, then cooled to room temperature. The absorbance at 760 nm was recorded as a measure of the total reducing capacity of the oil extract, which was compared with a calibration curve for gallic acid (Aldrich, Milwaukee, WI, USA) at different concentrations in 500 μL of methanol/water. Gallic acid is used widely as a standard for calibration in the Folin-Ciocalteu assay(*3*). The result for each oil extract was expressed as gallic acid equivalents of the 500 μL of methanol/water mixture used in the assay (i.e. the oil extracts were diluted 1:4 and this diluted solution was compared with the gallic acid calibration curve). This was necessary because 500 μL of the undiluted methanol/water extract saturated the reagent and produced a result which was outside the linear range of the assay.

As well as olive oil extracts, individual phenolic compounds were analysed using the Folin-Ciocalteu assay. Each standard phenolic compound was analysed at different concentrations in a 500 μL methanol/water mixture just as gallic acid was. A concentration-response curve for each compound was established, with the slope of the linear regression line revealing the relative

strength of response of the assay to each phenolic compound. Responses of standard phenolic compounds were based on linear least-squares fits of absorbance against concentration for at least six different concentrations. The responses of these compounds were compared with their $IC_{50}$ values (the concentration necessary to inhibit 50 % of radical attack) in the SIFT-MS-TOSC assay to determine the importance of each compound in the results of each assay.

### 9.2.2 HPLC Phenolic Analysis Method

The aim of this investigation was to evaluate the concentrations of each of the selected standard phenolic compounds in the olive oil samples and determine how the concentrations of these compounds related to the response observed in either the Folin-Ciocalteu or SIFT-MS-TOSC assays. The extraction of olive oil phenolic compounds followed the method of Montedoro et. al.(*4*) as adapted by Tovar et. al.(*1*). 20 mL of methanol/water 80:20 v/v was mixed with 45 g olive oil and separated by centrifugation at 1900 rpm for 10 min. Another 20 mL of methanol/water was mixed with the separated oil and also separated by centrifugation. The two extracts were combined to give 40 mL of methanol/water extract. This was concentrated in a rotary evaporator at 40 °C until a syrup-like consistency was reached. This was dissolved in 5 mL acetonitrile and washed with 20 mL hexane three times. The total 60 mL of hexane was washed with a further 5 mL acetonitrile. The resulting 10 mL acetonitrile was evaporated in a rotary evaporator and this syrup-like mixture dissolved in 5 mL acetonitrile. 2 mL of this solution was evaporated and the resulting syrup-like mixture dissolved in 1 mL methanol. The steps here which were omitted from the Folin-Ciocalteu assay procedure

were necessary for HPLC analysis as both a higher concentration and purity were demanded. This final 1 mL methanol solution is subsequently referred to as 'the extract' for all work concerning HPLC analysis throughout this chapter.

The separation and detection of extract components by High Performance Liquid Chromatography (HPLC) followed the method of Tovar et. al.(*1*), which is a slightly faster version of that used by Brenes et. al.(*5*). The HPLC system consisted of a Dionex ASI-100 autosampler, P680 HPLC pump, TCC-100 thermostatted column compartment, UVD340U detector and an Alltech ELSD800 evaporative light scattering detector. The column was a Waters Spherisorb ODS-2 (5 µm, 25 cm x 4.6 mm i.d., Alltech Associates, Inc., Deerfield, IL, USA) maintained at 35 °C. The mobile phases were 0.2 % acetic acid in water and methanol, flowing at 1 mL $min^{-1}$. The initial composition was 90 % water and 10 % methanol. This was changed to 30 % methanol in 10 min and kept at 30 % for 15 min. Methanol was raised to 40 % in 10 min and maintained for 5 min. It was then raised to 50, 60, 70 and 100 % in 5 min intervals. Initial conditions were reached in 15 min to give a total run time of 75 min. Chromatograms were obtained at 280 nm.

Standards of important, commercially available olive oil antioxidant compounds were used to obtain identification and quantification from HPLC analysis. The concentrations of the antioxidant compounds in the extracts were determined through the use of gallic acid (Aldrich, Milwaukee, WI, USA) as an internal standard, as none of the olive oils analysed contained detectable concentrations of gallic acid. Gallic acid provides a convenient internal standard because it is similar enough in structure to the other phenolic compounds to display a similar retention time and UV absorption intensity at the chosen

detection wavelength, yet was not present in the olive oil samples at detectable concentrations. The standard phenolic compounds used were tyrosol (Maybridge, Trevillett, Tintagel, Cornwall, UK), hydroxytyrosol (Tokyo Kasei Kogyo Co., Ltd., Tokyo, Japan), trolox (Aldrich, Milwaukee, WI, USA), p-coumaric acid (Fluka, Buchs, Switzerland) and vanillic acid (Sigma, St. Louis, MO, USA).

### 9.2.3 SIFT-MS-TOSC Assay

The antioxidant capacities of twenty olive oils were analysed using the peroxyl radical in the SIFT-MS-TOSC assay. Some standards were also prepared of known important olive oil antioxidant species (all of which were phenolic compounds) and analysed using the SIFT-MS-TOSC assay to obtain information on the relationships between the presence of different functional groups and antioxidant capacity. These antioxidant results were compared with the same standard phenolic compounds in the Folin-Ciocalteu assay, where the mechanism is known to involve electron transfer from phenolate ions (phenolic hydroxyl groups missing the acidic proton(*6*)).

Each standard phenolic compound (e.g. tyrosol) was dissolved in surfactant solution to give a 2 mmol $L^{-1}$ solution of the standard. This solution was diluted by differing the volumes to give solutions covering a range of concentrations. 600 μL of each diluted standard in surfactant solution was added to 3 mL of emulsion prepared with refined (‘Extra Light’) olive oil and mixed with a vortex mixer. The refined ‘Extra Light’ oil had been previously shown to have a very low antioxidant capacity. 600 μL of this emulsion mixture was added to the final assay bottles to obtain the same volume of oil emulsion as was used to analyse the olive oil samples. In this way, the only difference from

the routine method was an extra 100 µL surfactant solution, which was compensated for by adding 100 µL less buffer than in the routine method (3300 µL as opposed to 3400 µL). To cancel out any inhibition due to the olive oil used, the control also contained the emulsion, made up in the same way as for the standards only with no standard added. For example, in the case of the gallic acid standard the altered emulsion mixture was composed as shown in table 9.1 (with volumes in µL).

**Table 9.1.** Volumes (in µL) of surfactant, phenolic standard solution and emulsion used to produce the refined olive oil emulsion spiked with gallic acid for the SIFT-MS-TOSC assay.

| Bottle no. | Surfactant solution | Gallic acid solution | Emulsion |
|---|---|---|---|
| 1 | 600 | 0 | 3000 |
| 2 & 3 | 450 | 150 | 3000 |
| 4 & 5 | 300 | 300 | 3000 |
| 6 & 7 | 150 | 450 | 3000 |

This produced a control and 83.3, 167 and 250 µmol $L^{-1}$ solutions of gallic acid in the emulsion mixture. 600 µL diluted to 10 mL in the final assay bottles gave 5, 10 and 15 µmol $L^{-1}$ solutions of gallic acid respectively. The inhibition of ethene production caused by each phenolic compound was measured for at least six different concentrations which spanned a range of inhibitions covering the range from 30 % to 70 % inhibition. A 70 % inhibition corresponds to a 70 % reduction in ethene production compared to the control.

Six standard phenolic compounds were obtained: p-coumaric acid, gallic acid, vanillic acid, trolox, tyrosol and hydroxytyrosol. The first four of these compounds possess carboxylic acid groups, while the last two do not (figure 9.1). Each standard compound was analysed at different concentrations

to establish a response curve – the same method as is commonly used for gallic acid calibration in the Folin-Ciocalteu assay. This allows direct comparison of the relative responses of the antioxidant compounds in the two assays. By considering the structures of the compounds, information on the mechanisms of the assays may be gained. This information may be useful for explaining how different assays can give different results. It can also give an indication of which assay is the most relevant to use in a particular situation.

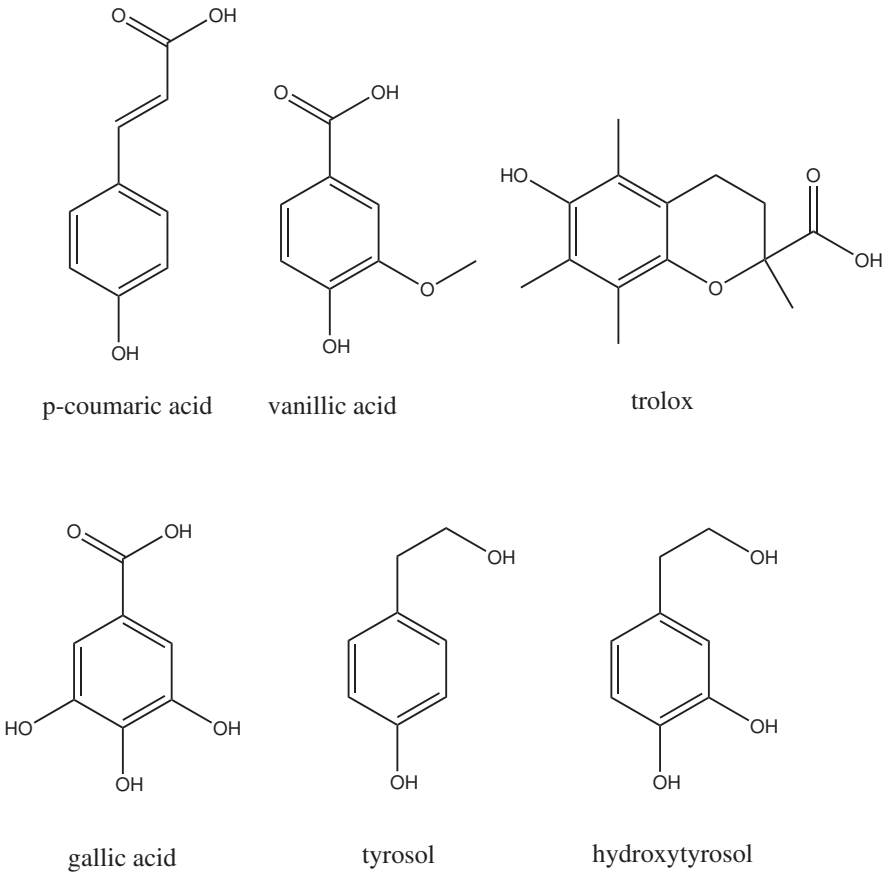



**Figure 9.1.** The six phenolic compounds studied. Neither tyrosol nor hydroxytyrosol contain a carboxylic acid group. Note that only hydroxytyrosol and gallic acid contain more than one phenolic hydroxyl group, having two and three respectively.

## 9.3 Results

### 9.3.1 SIFT-MS-TOSC Assay

When olive oil samples were analysed using the SIFT-MS-TOSC assay, it was found that results from the hydroxyl radical assay were not as satisfactory as those of the peroxyl radical. An unacceptable lack of precision was discovered during all single concentration hydroxyl radical assays attempted. The underlying reasons for this are not obvious. The intra-sample variation was greater than the inter-sample variation across all samples analysed, making the hydroxyl radical assay appear to be a pointless exercise. This variation was not apparent during the six point assays conducted earlier; neither the initial assays nor the single concentration calibration investigation suggested a problem might exist.

The hydroxyl radical assay results were investigated in order to discover the source of the large variation. The hydroxyl radical reference function (figure 8.4) is less steep than the peroxyl radical reference function (figure 8.3) for a large proportion of inhibition values. This increases the variation in $VI_{50}$ values derived from inhibition values for the hydroxyl radical assay much more than for the peroxyl radical assay. Further, the average head space ethene concentrations measured for the hydroxyl radical assay were lower than those measured for the peroxyl radical assay. This increases the relative error expected for the hydroxyl radical results. Factoring in both of these sources of error, 7.3 % and 28.3 % variation is expected for the peroxyl and the hydroxyl radical assays respectively. Both of these are slightly higher than the observed values of 6.6 % and 21.6 % respectively, therefore the maximum measured concentration and the slope of the reference function are together considered to

account for the poor performance of the hydroxyl radical assay. The maximum measured concentration of ethene may be increased, however the reference function is a characteristic of the assay and its slope cannot be altered. Only the peroxyl radical assay was used in subsequent analyses.

Results for all 20 olive oils analysed by the peroxyl SIFT-MS-TOSC assay are shown in table 9.2. These will be discussed in more detail in conjunction with the Folin-Ciocalteu assay results. Results for the standard phenolic compounds are shown in table 9.3. The four acids all gave similar responses, while tyrosol required more than six times the concentration of the acids to reach 50 % inhibition. Hydroxytyrosol was the most powerful of all, requiring approximately one third of the concentration of the acidic substances used to achieve 50 % inhibition.

**Table 9.2.** Peroxyl radical SIFT-MS-TOSC assay results (volumes in μL required to reduce the ethene concentration to half that found in the control, or $VI_{50}$) for the 20 olive oils included in this study. The means, standard deviations and coefficients of variation (CVs, standard deviations expressed as percentages of the means) of three single-point measurements of the $VI_{50}$ values are shown. The mean of the CV values is provided at the bottom of the table.

| | $VI_{50}$ | Std Dev | CV (%) |
|---|---|---|---|
| Oil 1 | 30.6 | 2.4 | 7.7 |
| Oil 2 | 47.9 | 10 | 21 |
| Oil 3 | 23.7 | 0.38 | 1.6 |
| Oil 4 | 48.1 | 2.4 | 5 |
| Oil 5 | 102 | 5.5 | 5.4 |
| Oil 6 | 81.8 | 9.2 | 11 |
| Oil 7 | 82.7 | 6.1 | 7.4 |
| Oil 8 | 88.8 | 7.2 | 8.1 |
| Oil 9 | 54.3 | 0.8 | 1.5 |
| Oil 10 | 52.7 | 4.4 | 8.4 |
| Oil 11 | 49.9 | 2.5 | 5 |
| Oil 12 | 34.6 | 0.94 | 2.7 |
| Oil 13 | 24.2 | 0.85 | 3.5 |
| Oil 14 | 32.2 | 0.29 | 0.92 |
| Oil 15 | 51.5 | 4.9 | 9.6 |
| Oil 16 | 126 | 20 | 16 |
| Oil 17 | 42.9 | 2 | 4.6 |
| Oil 18 | 37.4 | 2.2 | 5.8 |
| Oil 19 | 35.2 | 2.3 | 6.6 |
| Oil 20 | 233 | 0.91 | 0.39 |
| | CV Mean | | 6.6 |

**Table 9.3.** Curve fitting parameters for the pure phenolic compounds used in the SIFT-MS-TOSC assay. The equation used to fit the curves was $y = a\left(1\text{-}e^{bx}\right)$, the same as that of the emulsified oil reference function. The $R^2$ value for the fit is given in each case, along with the calculated concentration ($\mu mol\ L^{-1}$) required to give 50 % inhibition ($IC_{50}$).

| | a | b | $R^2$ | $IC_{50}$ |
|---|---|---|---|---|
| gallic acid | 187 | -0.0333 | 0.987 | 9.33 |
| p-coumaric acid | 100 | -0.0768 | 0.986 | 8.99 |
| tyrosol | 74.2 | -0.0178 | 0.995 | 63 |
| hydroxytyrosol | 95.2 | -0.232 | 0.872 | 3.22 |
| vanillic acid | 111 | -0.0791 | 0.991 | 7.56 |
| trolox | 161 | -0.0408 | 0.992 | 9.12 |

The results in table 9.3 suggest that the response observed in the SIFT-MS-TOSC assay does not rely heavily on the number of phenolic hydroxyl groups a compound possesses. This is demonstrated by gallic acid, which has three phenolic hydroxyl groups yet showed a similar response to p-coumaric acid, vanillic acid and trolox, all of which only have one phenolic hydroxyl group. Conversely, the response observed in the Folin-Ciocalteu assay is strongly governed by the number of phenolic hydroxyl groups, as reported by Singleton et. al.(*3*). Lichtenthäler et. al.(*7*) employing the TOSC assay, analysed a different set of compounds from those chosen for the present research also containing different numbers of phenolic hydroxyl groups, yet their results agree well with those of Singleton et. al. Catechin and epicatechin, each with four phenolic hydroxyl groups, were found to display the greatest inhibition against the peroxyl radical, whereas compounds containing fewer phenolic hydroxyl groups (such as trolox and protocatechuic acid, with one and two respectively, figure 9.2) did not rate as highly. The SIFT-MS-TOSC assay result for trolox obtained during the present research was not far removed from that of

Lichtenthäler et. al., however this was the only pure compound included in both studies, as Lichtenthäler et. al. did not investigate olive oil.

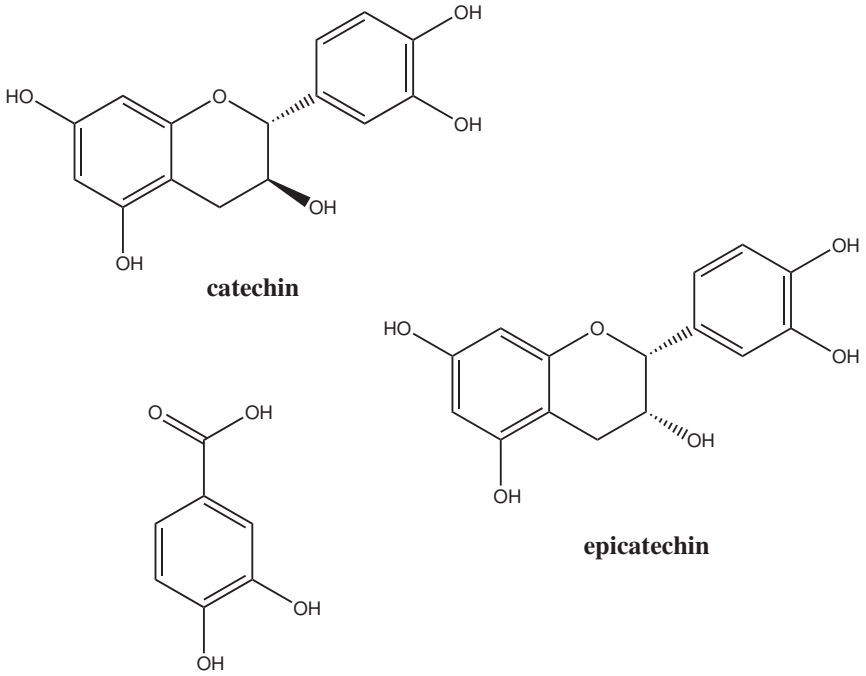



**Figure 9.2.** Structures of phenolic standard compounds studied by Lichtenthäler et. al.(*7*) using the TOSC assay.

If the phenolic hydroxyl groups were important for radical scavenging under SIFT-MS-TOSC assay conditions, p-coumaric acid would be expected to give a much lower response than gallic acid, while being approximately the same as that of tyrosol. This is because p-coumaric acid and tyrosol both have only one phenolic hydroxyl group, while gallic acid has three (figure 9.1). It is clear from table 9.3 that this is not the case, suggesting that an aspect of the molecule other than (or possibly in addition to) the number of phenolic hydroxyl groups it possesses is governing its radical scavenging capacity. It could be hypothesised that the carboxylic acid group was governing the assay response if it was not for the exceptionally low $IC_{50}$ of hydroxytyrosol. The

present results show a definite difference between the importance of reacting groups involved in the Folin-Ciocalteu and SIFT-MS-TOSC assays.

An interesting trend was noticed in the results obtained here. As for the oils analysed using the peroxyl radical assay in section 8.2.2, the concentrations of the standard phenolic compounds were divided by their $IC_{50}$ values to produce a function that reached 50 % inhibition at the arbitrary concentration value of one. When scaled in this way, all standard phenolic compounds followed the same curve. In fact, the phenolic compounds followed the same curve as the oils analysed in section 8.2.2 (figure 9.2). The $R^2$ value for the oils following their best fit equation (equation 9.1) is 0.97, while for the standard phenolic compounds it is 0.96. These results suggest that equation 9.1 is characteristic of the SIFT-MS-TOSC peroxyl radical assay, and can be used to describe any antioxidant species. This knowledge allows the use of a single-point assay for any emulsion-based antioxidant sample, and supports the hypothesis from section 8.2.2 that the shape of the inhibition-volume curve is determined by a simple probability relationship and not by the nature of the antioxidants present. Further investigation is necessary before applying equation 9.1 to aqueous samples, yet it is anticipated these will also be well described.

$$y = 97.65\left(1-e^{-0.7178x}\right) \qquad (9.1)$$

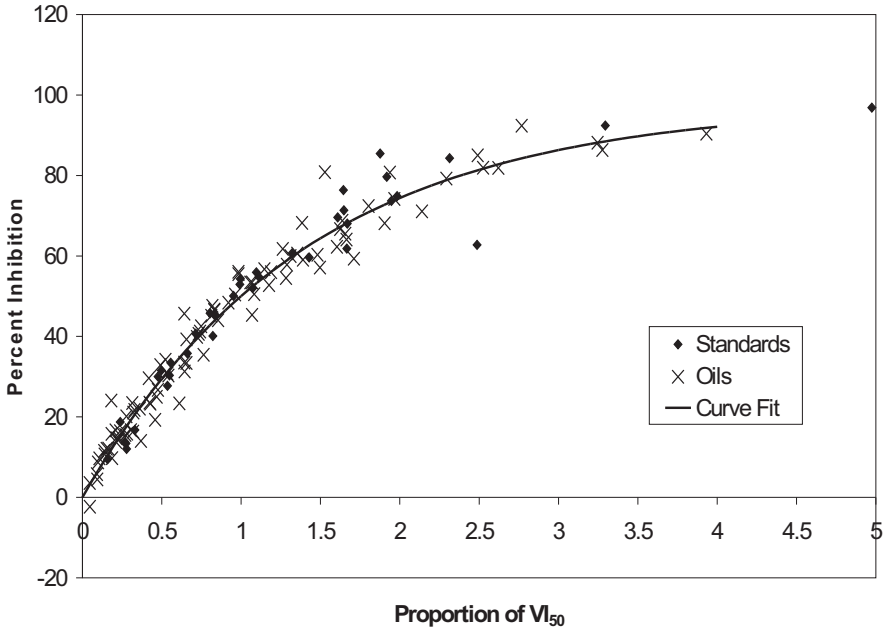



**Figure 9.2.** Percent inhibition of radical attack produced by standard phenolic compounds and oils at different concentrations relative to their $VI_{50}$ values, analysed in emulsion by the SIFT-MS-TOSC peroxyl radical assay. The curve fit follows equation 9.1, which is a good approximation for all samples yet analysed.

### 9.3.2 Folin-Ciocalteu Assay

The assay procedure as outlined in Scalbert et. al.(*2*) required little alteration to provide an absorbance for the samples which was within the linear range of the calibration curve obtained for gallic acid. The 1:4 dilution given in the method above was sufficient for all but one of the samples (table 9.4). Oil 20, which showed by far the lowest response in the SIFT-MS-TOSC assay, gave a value below the linear range of the gallic acid response curve. The variation in the absorbance for the oil samples gave only a 4 % standard deviation between calculated values, which was less than that of the SIFT-MS-TOSC assay. However, as this assay is based on a linear calibration which allowed the errors

in the slope and intercept to be taken into account, more realistic error estimates were calculated to obtain the 15 % CV given below. This should not be interpreted as reporting a larger variation from the Folin-Ciocalteu assay than from the SIFT-MS-TOSC assay.

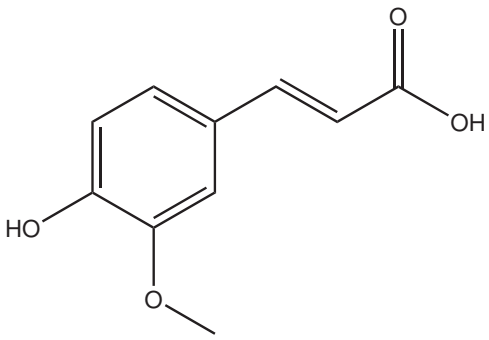
**Table 9.4.** Folin-Ciocalteu assay results (mg $L^{-1}$ gallic acid equivalents, GAE) for the 20 olive oils included in this study. The means, standard deviations and coefficients of variation (CVs, standard deviations expressed as percentages of the means) of three measurements of the GAE concentration are shown. The mean of the CV values is provided at the bottom of the table. Curve-fitting parameters for the gallic acid calibration function are included in table 9.5.

| | GAE | Std Dev | CV (%) |
|---|---|---|---|
| Oil 1 | 80 | 10 | 12 |
| Oil 2 | 53 | 7 | 13 |
| Oil 3 | 88 | 9 | 10 |
| Oil 4 | 66 | 8 | 12 |
| Oil 5 | 31 | 4 | 13 |
| Oil 6 | 33 | 5 | 14 |
| Oil 7 | 26 | 5 | 18 |
| Oil 8 | 27 | 3 | 13 |
| Oil 9 | 26 | 3 | 12 |
| Oil 10 | 34 | 5 | 14 |
| Oil 11 | 42 | 8 | 19 |
| Oil 12 | 73 | 8 | 11 |
| Oil 13 | 110 | 10 | 13 |
| Oil 14 | 80 | 8 | 10 |
| Oil 15 | 52 | 6 | 11 |
| Oil 16 | 9 | 4 | 47 |
| Oil 17 | 55 | 7 | 13 |
| Oil 18 | 52 | 8 | 15 |
| Oil 19 | 53 | 7 | 14 |
| Oil 20 | Below linear range | | |
| | Mean CV | | 15 |

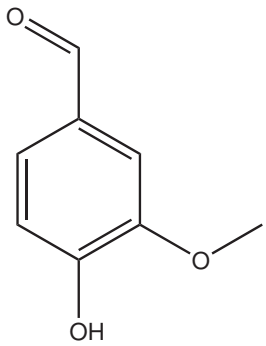
The size of the linear range is dependent both on the concentration of the Folin-Ciocalteu reactive species available to react and on the absorbance range able to be measured by the spectrometer. The response obtained from a sample

is expressed as gallic acid equivalent concentration, and has no meaning other than to relate samples to each other. Samples may be diluted as much as desired (provided they still give an assay response within the linear range of the gallic acid calibration curve), but all samples must be compared at the same level of dilution.

Results for the pure phenolic compounds are shown in table 9.5. Results for only gallic acid and p-coumaric acid were available in the literature(*3*), and while the relative results for the two compounds determined in this study agree well, the absolute values for molar absorption are different. This may be due to a difference in absorption scale between the two spectrometers used, or perhaps the value from Singleton et. al.(*3*) was provided in different units. As no units were given, this is unknown. Only two of the phenolic hydroxyl groups of gallic acid are reported to react(*3*), therefore its molar absorption is expected (and was observed) to be similar to that of hydroxytyrosol. All of the other compounds have one phenolic hydroxyl group and are expected to show similar molar absorption values. This was observed, with one exception. Vanillic acid gave a result approximately seven to eight times lower than expected. The methoxy group of vanillic acid is not considered to cause the low response, however vanillic acid possesses few other features which could have caused this response. Singleton et. al.(*3*) studied a list of compounds which included ferulic acid and vanillin, both of which have methoxy groups (figure 9.3). All compounds studied by Singleton et. al. gave the expected responses, based on the number of phenolic hydroxyl groups they possess.




**ferulic acid**




**vanillin**

**Figure 9.3.** The structures of ferulic acid and vanillin, both similar to vanillic acid in that they each possess a single phenolic hydroxyl and methoxy group.

**Table 9.5.** Values obtained from Folin-Ciocalteu assay investigations of standard phenolic compounds. Absorbance was modelled by the equation for a straight line $A = mx + c$, where m and c are the constants listed in the table and x is the concentration in mmol $L^{-1}$. The $R^2$ value of the regression line is given, as is the molar absorption coefficient ε (= m/l where m is the absorbance slope in mol $L^{-1}$ and l is the path length in cm, using Beer's law, $A = \varepsilon x l$). The standard error of the molar absorption is also given, calculated from the fit of the regression line (as described in section B.3). The molar absorbance from Singleton et. al.(*3*) is also shown for those compounds included in both investigations, however no units were supplied by Singleton et. al. so it is not clear whether the values are directly comparable.

| | m | c | $R^2$ | ε ($L\ mol^{-1}\ cm^{-1}$) | ε standard error | Molar absorbance |
|---|---|---|---|---|---|---|
| gallic acid | 1.56 | 0.111 | 0.991 | 1560 | 70 | 25000 |
| p-coumaric acid | 0.951 | 0.126 | 0.995 | 950 | 30 | 15600 |
| tyrosol | 0.812 | 0.0579 | 0.997 | 810 | 20 | |
| hydroxytyrosol | 1.6 | 0.0309 | 0.983 | 1600 | 90 | |
| vanillic acid | 0.102 | 0.0209 | 0.973 | 100 | 10 | |
| trolox | 0.65 | 0.0203 | 0.992 | 650 | 20 | |

The possibility of decomposition of phenolic compounds was considered for vanillic acid from the Folin-Ciocalteu results and tyrosol from the SIFT-MS-TOSC results, as this is the most simple explanation for the low responses observed. No pattern linking the functional groups present with the results was discovered that would account for the low responses of these two compounds in the respective assays. Decomposition of the compounds was not considered likely for several reasons. The most prominent reason was the lack of agreement between the assays. If one compound had decomposed and given a low response in one assay, it would also be expected to give a low response in the other assay. Of course this assumes that any break down products give a negligible response in both assays. The HPLC results also suggested there was no break down occurring, as the relative retention times observed were consistent with those of Tovar et. al.(*1*), the two calibration standards used for tyrosol both produced identical UV/Vis spectra, and no erroneous peaks were observed from the ELSD results of calibration mixtures (meaning that no other compounds, whether they possessed UV active chromophores or not were present in the calibration samples). The low responses of the two compounds in their respective assays is interesting and may warrant further investigation in the future.

### 9.3.3 Comparison of Two Assays

A comparison of the results for the olive oil samples in the Folin-Ciocalteu and SIFT-MS-TOSC assays are presented in figure 9.4. An inverse relationship makes sense due to the differences in the two assays. If a sample were to give a zero result for the Folin-Ciocalteu assay, it would have an

infinite 50 % inhibition volume in the SIFT-MS-TOSC assay. Zero values in the SIFT-MS-TOSC assay and infinite values in the Folin-Ciocalteu assay are impossible, so asymptotes are conceptually valid when relating these two assays to each other. The relationship between the results was significant, yet there was still noticeable scatter away from the line of best fit. This result was exactly as expected, as the Folin-Ciocalteu assay measures only phenolic compounds which are soluble in a mixture of methanol and water, while the SIFT-MS-TOSC assay measures all antioxidant compounds in a sample. Some antioxidants are hydrophobic and cannot be analysed by the Folin-Ciocalteu assay, as they are not present in the oil extract. Such compounds may include tocopherols, tocotrienols and other related compounds(*8*).

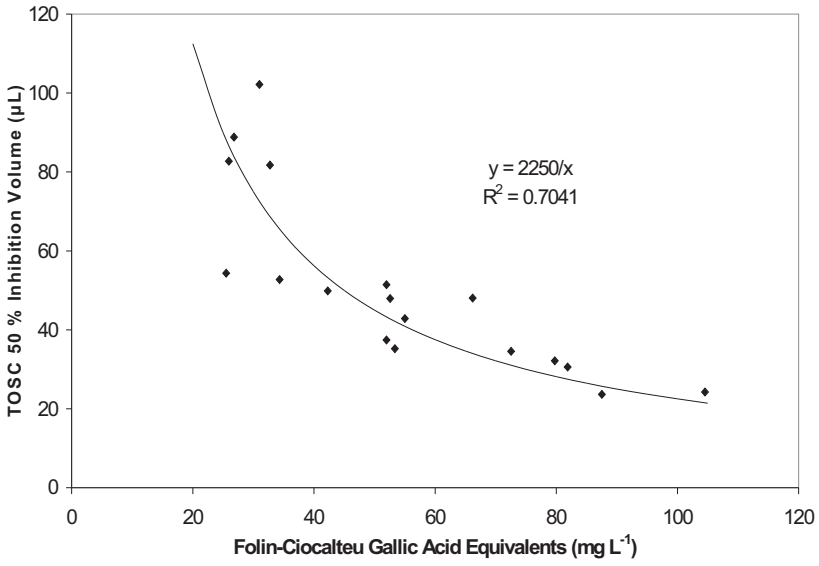



**Figure 9.4.** Responses of 18 olive oils in the peroxyl radical SIFT-MS-TOSC assay and Folin-Ciocalteu total phenols assay. Two outliers were removed due to their very low responses in the Folin-Ciocalteu assay. A significant relationship was observed, with an $R^2$ value of 0.70.

The present results suggest that polar phenolic compounds comprise the bulk of olive oil antioxidants because it is these compounds that are featured in the Folin-Ciocalteu assay. It is not known whether the majority of the deviation of the SIFT-MS-TOSC assay data from the best fit line is due to the differences in response of the different antioxidant species in each of the two assays or the contributions from hydrophobic antioxidants, although both factors must contribute to some extent.

### 9.3.4 High-Performance Liquid Chromatography Results

To evaluate the importance of the selected phenolic compounds to the response of olive oil in the two assays, the concentrations of the selected compounds were determined in extracts taken from the olive oil samples by High-Performance Liquid Chromatography (HPLC). Neither gallic acid nor trolox were identified in olive oil using the method of Tovar et. al.(*1*), so only the remaining four phenolic compounds shown in figure 9.1 were measured, with gallic acid as an internal standard. Tyrosol and hydroxytyrosol are consistently named among the most highly concentrated phenolic compounds in olive oil, with vanillic acid and p-coumaric acid also commonly identified, yet at lower concentrations(*1;4;9;10*). Hrncirik and Fritsche(*10*), using a slight alteration of the extraction procedure employed here and Montedoro(*4*), using essentially the same procedure, both extracted over 90 % of olive oil phenolic compounds with methanol/water mixtures. Based on these findings, the extracts used in the present study were considered to represent well the phenolic components of the olive oils from which they came.

The phenolic standards were evaluated in triplicate at six different concentrations ranging from 2.5 to 1000 μg $mL^{-1}$ against 1 mg $mL^{-1}$ gallic acid.

The concentration range spanned for calibration depended on the concentration of each phenolic compound observed in the extract. The retention times and response parameters are displayed in table 9.6.

**Table 9.6.** Retention times (in minutes) and response parameters for four phenolic compounds measured by HPLC against 1 mg $mL^{-1}$ gallic acid. Parameters are for linear calibration functions y = mx + c, where y = area of compound peak / area of gallic acid peak and x = mg $mL^{-1}$ of compound in the calibration solution. Each calibration standard was analysed in triplicate. The $R^2$ value of each calibration fit is given, as is the standard error ($\sigma_m$) of the slope of the regression line. The limit of detection (LOD, in μg per mL of extract) is also included, based on the lowest concentration calibration sample in which each compound was able to be detected.

| | RT | m | $\sigma_m$ | c | $R^2$ | LOD |
|---|---|---|---|---|---|---|
| hydroxytyrosol | 7.87 | 0.36 | 0.05 | -0.008 | 0.997 | <25 |
| tyrosol | 10.7 | 0.212 | 0.001 | -0.0048 | 0.999 | 50 |
| vanillic acid | 13.7 | 0.615 | 0.008 | -0.002 | 0.997 | 20 |
| p-coumaric acid | 18.6 | 1.6 | 0.02 | 0.0072 | 0.996 | <2.5 |

The concentrations of the four phenolic compounds analysed in the twenty olive oil samples are shown in tables 9.7 and 9.8. Tyrosol was generally present at the highest concentration, followed by hydroxytyrosol. Vanillic acid and p-coumaric acid were only found at low concentrations in the extracts.

**Table 9.7.** Concentrations of hydroxytyrosol and tyrosol found in olive oil extracts by HPLC. Each oil was analysed in duplicate and quantified by a five point linear calibration (table 9.6) using gallic acid as an internal standard.

| | µg per mL extract | | | |
|---|---|---|---|---|
| | hydroxytyrosol | Std Dev | tyrosol | Std Dev |
| Oil 1 | 190 | 10 | 180 | 20 |
| Oil 2 | 500 | 100 | 1400 | 100 |
| Oil 3 | 920 | 30 | 650 | 10 |
| Oil 4 | 110 | 20 | 450 | 10 |
| Oil 5 | 90 | 20 | 900 | 100 |
| Oil 6 | 0 | 0 | 140 | 10 |
| Oil 7 | 0 | 0 | 130 | 10 |
| Oil 8 | 0 | 0 | 810 | 20 |
| Oil 9 | 80 | 20 | 400 | 30 |
| Oil 10 | 125 | 9 | 500 | 30 |
| Oil 11 | 0 | 0 | 190 | 20 |
| Oil 12 | 220 | 30 | 740 | 50 |
| Oil 13 | 190 | 10 | 224 | 5 |
| Oil 14 | 130 | 20 | 257 | 8 |
| Oil 15 | 50 | 7 | 90 | 20 |
| Oil 16 | 0 | 0 | 189 | 10 |
| Oil 17 | 150 | 20 | 520 | 10 |
| Oil 18 | 240 | 30 | 450 | 20 |
| Oil 19 | 150 | 20 | 320 | 10 |
| Oil 20 | 0 | 0 | 0 | 0 |

**Table 9.8.** Concentrations of vanillic acid and p-coumaric acid found in olive oil extracts by HPLC. Each oil was analysed in duplicate and quantified by a five point linear calibration using gallic acid as an internal standard. The concentrations of these phenolic compounds found in olive oil were much lower than those of tyrosol and hydroxytyrosol.

| | μg per mL extract | | | |
|---|---|---|---|---|
| | vanillic acid | Std Dev | p-coumaric acid | Std Dev |
| Oil 1 | 0 | 0 | 0 | 0 |
| Oil 2 | 0 | 0 | 0 | 0 |
| Oil 3 | 0 | 0 | 0 | 0 |
| Oil 4 | 0 | 0 | 0 | 0 |
| Oil 5 | 20 | 30 | 5 | 2 |
| Oil 6 | 0 | 0 | 0 | 0 |
| Oil 7 | 0 | 0 | 0 | 0 |
| Oil 8 | 0 | 0 | 0 | 0 |
| Oil 9 | 30 | 50 | 10 | 2 |
| Oil 10 | 0 | 0 | 10 | 20 |
| Oil 11 | 0 | 0 | 0 | 0 |
| Oil 12 | 50 | 4 | 0 | 0 |
| Oil 13 | 0 | 0 | 4 | 3 |
| Oil 14 | 0 | 0 | 0 | 0 |
| Oil 15 | 0 | 0 | 0 | 0 |
| Oil 16 | 0 | 0 | 0 | 0 |
| Oil 17 | 21 | 4 | 0 | 0 |
| Oil 18 | 0 | 0 | 0 | 0 |
| Oil 19 | 0 | 0 | 0 | 0 |
| Oil 20 | 0 | 0 | 0 | 0 |

Principal Component Analysis (PCA) was performed on the results obtained for the 18 oils involved in figure 9.4. The SIFT-MS-TOSC assay $VI_{50}$ values (made linear by raising each $VI_{50}$ value to power of -1), the Folin-Ciocalteu gallic acid equivalent values, the extract concentrations of the four phenolic compounds and the total HPLC peak area (minus the peak area of the gallic acid internal standard) were standardised (mean-centred and made to have standard deviations of one) and formed into a **Z** matrix (see section A.1.1).

Four PCs were judged to be significant from assessment of their respective eigenvalues, however the first PC was by far the most interesting.

The PC loadings (**C**) matrix is shown in table 9.9, showing the correlations between the PCs and the results for 18 oils. PC 1 represents the largest variation found in the data, and has correlations of greater than 0.7 with the Folin-Ciocalteu, SIFT-MS-TOSC results and the hydroxytyrosol concentration. PC 1 also has a correlation of -0.61 with the total HPLC peak area, while the concentrations of tyrosol, vanillic acid and p-coumaric acid are independent, being represented by the other significant PCs.

**Table 9.9.** Principal component loadings (**C**) matrix from PCA performed on the concentrations of phenolic compounds determined in olive oil extracts, peroxyl radical SIFT-MS-TOSC assay $VI_{50}$ values, Folin-Ciocalteu gallic acid equivalents and the total peak area of the olive oil extracts obtained from HPLC with the area of the internal standard peak subtracted. PC 1 shows significant correlations with hydroxytyrosol, 1 / (SIFT-MS-TOSC $VI_{50}$), Folin-Ciocalteu gallic acid equivalents and total HPLC peak area. PCs 2, 3 and 4 show significant correlations with tyrosol and vanillic acid (PC 2), tyrosol (PC 3) and p-coumaric acid (PC 4). Correlations above 0.60 were considered significant.

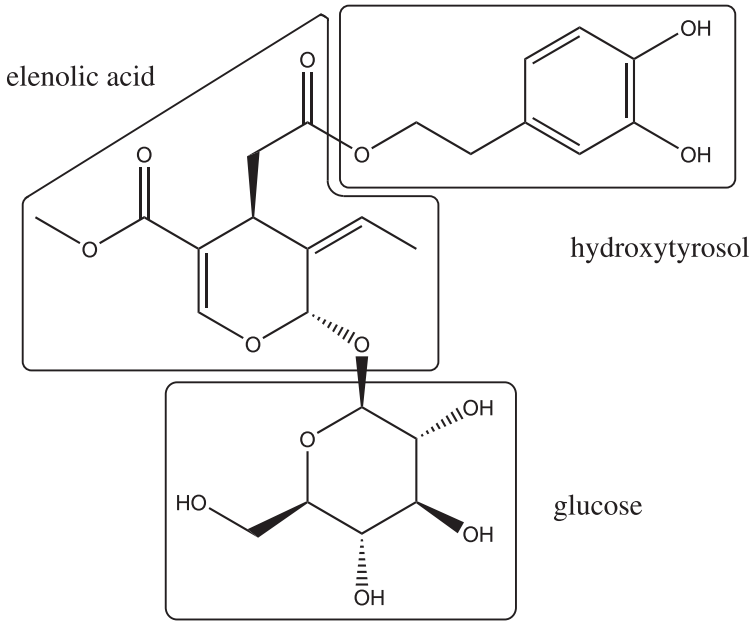
| **C** | | | | |
|---|---|---|---|---|
| | PC 1 | PC 2 | PC 3 | PC 4 |
| hydroxytyrosol | -0.751 | 0.128 | 0.557 | -0.0293 |
| tyrosol | -0.146 | 0.722 | 0.603 | 0.158 |
| vanillic acid | 0.0928 | 0.819 | -0.397 | -0.176 |
| p-coumaric acid | 0.283 | 0.0155 | 0.202 | -0.929 |
| 1 / (SIFT-MS-TOSC $VI_{50}$) | -0.918 | -0.245 | -0.0693 | -0.249 |
| Folin-Ciocalteu | -0.924 | -0.251 | -0.164 | -0.031 |
| total HPLC peak area | -0.612 | 0.549 | -0.441 | -0.038 |

PC 1 appears to represent an approximation of the antioxidant capacity of the samples, as it shows significant correlations with the two assays used to approximate this quantity (the Folin-Ciocalteu and SIFT-MS-TOSC assays). As

these correlations are negative, PC 1 in facts runs antiparallel to the antioxidant capacity, however the effect remains the same. The total HPLC peak area is expected to correlate with the antioxidant capacity, as the Folin-Ciocalteu assay results do, and both the HPLC and Folin-Ciocalteu analyses were performed on extracts which contained exclusively olive oil phenolic compounds (major contributors to olive oil antioxidant capacity). Hydroxytyrosol was identified in most of the oil samples and showed a high response in both the SIFT-MS-TOSC and Folin-Ciocalteu assays. This suggests that the hydroxytyrosol concentration had a strong influence over the antioxidant capacity, as reflected in its correlations with the SIFT-MS-TOSC assay (1 / these results to make them linear as performed earlier), Folin-Ciocalteu assay and the total HPLC peak area results (correlations of 0.94, 0.77 and 0.47 respectively, analysis not shown). The first two correlations are significant to a 99 % confidence level, while the last is significant to a 95 % confidence level(*11*).

There is not enough evidence here to prove that the hydroxytyrosol concentration is causing the correlations between the different tests, as there may be other compounds whose concentrations are linked with that of hydroxytyrosol and which have a larger effect on the antioxidant capacity. One potential example is oleuropein(*4;9*). Oleuropein is an ester in which elenolic acid (which includes a glucose molecule added at a different site) is linked to hydroxytyrosol (figure 9.5). This compound was measured at higher concentrations than hydroxytyrosol in many of the olive oil samples analysed by Montedoro et. al.(*4*), and also at high concentrations in its ‘aglycone’ form (without glucose) by Tovar et.al.(*1*). It is reasonable to assume that the linked and free forms of hydroxytyrosol (linked by condensation reactions, released by

hydrolysis reactions) are in equilibrium with each other, both in the oil and in the extract. The antioxidant capacity of oleuropein was not determined, yet if it is comparable with that of hydroxytyrosol (which seems to be a reasonable assumption), the antioxidant capacity of the olive oil samples analysed in the present research cannot be attributed to hydroxytyrosol alone.




**oleuropein**

**Figure 9.5.** The structure of oleuropein, a phenolic compound commonly found in olive oil. Oleuropein consists of glucose, elenolic acid and hydroxytyrosol components.

## 9.4 Summary

Comparison of the results for the individual phenolic compounds in the SIFT-MS-TOSC assay and the Folin-Ciocalteu assay did not uncover any direct patterns linking molecular structure with antioxidant capacity in the SIFT-MS-TOSC assay (table 9.10). Several structure-activity relationship studies have been performed using the TOSC assay(*7;12*), however the

compounds analysed were different from those investigated here and the results do not provide information useful in this study.

**Table 9.10.** Ranked order of responses of individual phenolic compounds in the SIFT-MS-TOSC and Folin-Ciocalteu antioxidant assays. The codes tyOH, va, pca, tx, ga and ty refer to hydroxytyrosol, vanillic acid, p-coumaric acid, trolox, gallic acid and tyrosol respectively. Double arrows indicate where one compound displayed more than twice the response of the following compound. Tables 9.3 and 9.5 contain the same information displayed in greater detail.

| |
|---|
| SIFT-MS-TOSC assay |
| tyOH >> va > pca > tx > ga >> ty |
| Folin-Ciocalteu assay |
| tyOH > ga > pca > ty > tx >> va |

The SIFT-MS-TOSC assay was performed on 20 olive oils and the results compared with those from both the Folin-Ciocalteu 'total phenols' assay and analysis of the phenolic compounds by HPLC with UV/VIS detection. Significant correlations were seen between all of these results, suggesting that the SIFT-MS-TOSC assay shows an acceptable amount of agreement with the widely accepted Folin-Ciocalteu assay and that both of these assays measure principally phenolic compounds in olive oil. These results are in line with many other researchers who have concluded that the phenolic compounds of olive oil make the major contribution to the antioxidant capacity. One reason for differences between the SIFT-MS-TOSC assay and the Folin-Ciocalteu assay results is hydrophobic antioxidants such as tocopherols which are measured only by the SIFT-MS-TOSC assay.

## 9.5 References


(1) Tovar, M. J.; Motilva, M. J.; Romero, M. P. Changes in the Phenolic Composition of Virgin Olive Oil from Young Trees (*Olea europaea* L. cv. Arbequina) Grown under Linear Irrigation Strategies. *J. Agric. Food Chem.* **2001**, *49*, 5502-5508.

(2) Scalbert, A.; Monties, B.; Janin, G. Tannins in Wood: Comparison of Different Estimation Methods. *J. Agric. Food Chem.* **1989**, *37*, 124-1329.

(3) Singleton, V. L.; Orthofer, R.; Lamuela-Raventós, R. M. Analysis of Total Phenols and Other Oxidation Substrates and Antioxidants by Means of Folin-Ciocalteu Reagent. *Methods Enzymol.* **1999**, *299*, 152-178.

(4) Montedoro, G.; Servili, M.; Baldioli, M.; Miniati, E. Simple and Hydrolyzable Phenolic Compounds in Virgin Olive Oil. 1. Their Extraction, Separation, and Quantitative and Semiquantitative Evaluation by HPLC. *J. Agric. Food Chem.* **1992**, *40*, 1571-1576.

(5) Brenes, M.; García, A.; García, P.; Ríos, J. J.; Garrido, A. Phenolic Compounds in Spanish Olive Oils. *J. Agric. Food Chem.* **1999**, *47*, 3535-3540.

(6) Huang, D.; Ou, B.; Prior, R. The Chemistry Behind Antioxidant Capacity Assays. *J. Agric. Food Chem.* **2005**, *53*, 1841-1856.

(7) Lichtenthäler, R.; Marx, F.; Kind, O. M. Determination of antioxidative capacities using an enhanced total oxidant scavenging capacity (TOSC) assay. *Eur. Food Res. Technol.* **2003**, *216*, 166-173.

(8) Huang, D.; Ou, B.; Hampsch-Woodill, M.; Flanagan, J. A.; Deemer, E. K. Development and Validation of Oxygen Radical Absorbance Capacity Assay for Lipophilic Antioxidants Using Randomly Methylated β-Cyclodextrin as the Solubility Enhancer. *J. Agric. Food Chem.* **2002**, *50*, 1815-1821.

(9) Boskou, D. *Olive Oil: Chemistry and Technology*; AOCS Press: Champaign, IL, USA, 1996.

(10) Hrncirik, K.; Fritsche, S. Comparability and reliability of different techniques for the determination of phenolic compounds in virgin olive oil. *Eur. J. Lipid Sci. Technol.* **2004**, *106*, 540-549.

(11) Chase, W.; Bown, F. *General Statistics*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1992.

(12) Dugas, A. J. J.; Castañeda-Acosta, J.; Bonin, G. C.; Price, K. L.; Fischer, N. H.; Winston, G. W. Evaluation of the Total Peroxyl Radical-Scavenging Capacity of Flavonoids: Structure-Activity Relationships. *J. Nat. Prod.* **2000**, *63*, 327-331.

# Chapter 10

## General Conclusions

### 10.1 Relationships Between Sensory Attributes and SIFT-MS VOC Analysis

#### 10.1.1 General Results

SIFT-MS provides a straightforward, rapid method of VOC analysis for the head space of olive oil. Thirteen analytes were identified using SIFT-MS and these could be quantified in a one minute scan by SIFT-MS.

Methanol and ethanol were identified as the highest concentration head space VOCs in the vast majority of olive oils and in olive pomace. This finding contradicts many literature reports of chromatographic analyses on olive oil volatile compounds and emphasises the advantages of SIFT-MS which provides a "snapshot" of the VOCs present without showing preference for any class of compounds. There are several reasons why chromatographic analysis may lead to the underestimation of polar compounds such as methanol and ethanol, mostly due to their high polarity when compared with larger organic molecules. These reasons include the use of sampling techniques such as Tenax and solid-phase microextraction which adsorb less polar compounds preferentially, columns designed to separate less polar compounds that show poor retention of methanol and ethanol and flame ionisation detection where larger molecules give a greater per-mole response than smaller ones. The use of an internal standard may underestimate the response of methanol and ethanol further if used to provide relative quantification of identified compounds. Any compound

larger than ethanol will produce a larger response per mole in a flame ionisation detector than methanol or ethanol, causing the relative concentrations of these small alcohols to appear less significant.

### 10.1.2 Unfavourable Sensory Attributes

The analysis of olive oil defects identified several characteristic volatile products of unfavourable processes which may occur in olive oil. An in-depth study into olive oil oxidation showed that propanal and acetone were the most prominent volatile products and could be used to predict the peroxide value of olive oils. The concentrations of peroxides and volatile compounds are not necessarily correlated, as correlations usually exist only if the oil has been oxidised under constant conditions (as in the study described here), and are thought to depend on storage temperature(*1*). This is why correlations between volatile compounds and sensory attributes were sought, as sensory measurements are based primarily on volatile compound concentrations. In order to fully assess the worth of the correlations established in this study, the study should be repeated at different temperatures (say 50 and 70 °C). However, oil is oxidised under constant conditions only in a research setting, so the results from the present studies would not be able to reliably predict the peroxide value for all oils. Instead, the results would be best used as an independent measure of oil oxidation, complementing the existing peroxide value and sensory evaluation tests. It may well eventuate that this test may be able to replace the sensory test, as they measure the same properties of the oil, only processing the data differently.

High concentrations of ethanol, acetic acid, ethyl acetate and at least one isomer of pentanol were found to be indicators of various types of olive oil

fermentation, and the concentrations of these VOCs were linked to the strengths of their respective defects. There are several fermentation defects that are caused by different bacteria. These are not easily distinguished using SIFT MS, however until they are investigated in greater depth, it is uncertain whether this is a significant limitation.

The musty defect, which arises from attack by mould, did not provide any characteristic products that could be identified by SIFT-MS analysis in this study. Moulds tend to produce products which are less volatile than those derived from other unfavourable processes, so are more difficult to detect(*2*). Musty olive oil may be able to be distinguished by a lack of important positive volatile compounds such as (E)-2-hexenal, however further investigation is necessary.

### 10.1.3 Favourable Sensory Attributes

The suitability of the SIFT-MS technique for the prediction of the favourable sensory attributes fruitiness, bitterness and pungency was assessed. No favourable attribute could be predicted from a consideration of SIFT-MS VOC analysis in the limited number of oils accessible in this work. Limitations in the data may well have contributed to the lack of correlation between the VOCs observed and the sensory attributes detected by the sensory panel. The attributes of bitterness and pungency are very much taste (as opposed to aroma) sensations, so are less likely to be assessed by a volatile-based test.

Enough information is obtained from a one minute SIFT-MS scan of olive oil head space to distinguish a low quality, defective oil from a high quality, fresh oil by visual inspection of individual volatile compound concentrations. Multivariate regression methods (such as partial least-squares

regression, and possibly some nonlinear methods) are likely to provide more information, such as the degree of oxidation described in chapter five. This level of characterisation may enable the SIFT-MS method to be used for rapid grading of olive oils. The replacement of the sensory test by an instrumental test such as this is still in the future, however SIFT-MS may be useful sooner in situations where sensory assessment is not possible, yet where a large number of olive oil samples are present to justify the investment. An import warehouse or a large oil production mill are the two most likely environments. The quality of olive oil could be rapidly assessed, ensuring that either an imported shipment of oil was not accepted if below a certain standard, or a freshly produced oil of low quality was not mixed with high quality oil, thereby preserving the standard of the high quality oil.

## 10.2 The SIFT-MS-TOSC Assay for Antioxidant Analysis

### 10.2.1 Further Assay Development

The SIFT-MS-TOSC assay is the only antioxidant assay at present which is able to assess both aqueous and emulsified samples with no change in assay conditions. This assay was used to evaluate antioxidant capacity (expressed as the volume of oil in μL necessary to inhibit ethene gas production by 50 % compared with a control; this is known as the $VI_{50}$ value) against two different radical species for a number of olive oil samples of different quality and origin. Reference functions were constructed for each radical which allowed the calculation of the $VI_{50}$ value of an oil from the inhibition observed at any single concentration. This development increased the sample turnover of the assay and simplified the assay preparation. The work involved in conducting the

SIFT-MS-TOSC assay remains significant, taking around four hours from start of preparation to completion, assuming the necessary pre-made solutions are available. This amount of work is not excessive for an antioxidant assay, so is acceptable at the present time. The ability to analyse two samples in the same amount of time using the inhibition-volume reference function is a noticeable improvement. The potential for automation remains an area for further improvement.

### 10.2.2 Comparison of Different Antioxidant Assays

Twenty olive oils were analysed by the SIFT-MS-TOSC assay, the widely used Folin-Ciocalteu total phenols assay and an HPLC method for quantifying selected phenolic compounds. Tyrosol and hydroxytyrosol were found in many of the olive oils at significant concentrations, but p-coumaric acid and vanillic acid were either not detected or detected at low concentrations. The two antioxidant assays and the total HPLC peak area were found to be significantly correlated with the same property of the olive oils. This property is most likely the true antioxidant capacity of the oils. As the Folin-Ciocalteu assay and the total HPLC peak area both measure the phenolic compounds of olive oil, the similarities between the assays confirm the reports that the phenolic compounds of olive oil are responsible for the bulk of the oil's antioxidant capacity.

Six different phenolic compounds were analysed by both the SIFT-MS-TOSC and Folin-Ciocalteu assays to compare their responses in the two assays. Hydroxytyrosol was found to be the most potent of the six phenolic compounds in both assays. It was found that phenolic hydroxyl groups react with the Folin-Ciocalteu reagent to give the response observed in the

Folin-Ciocalteu assay, however the important functional groups or molecular structures in the SIFT-MS-TOSC assay were not discovered.

The SIFT-MS-TOSC assay proved to be a reliable method for the analysis of antioxidant capacity in emulsified samples. The research described here did not involve significant work with aqueous samples, however the assay performed as reliably for the few aqueous samples analysed as for the emulsified samples. The peroxyl radical alone was used for the bulk of the research, as the hydroxyl radical results displayed low repeatability. Problems with the hydroxyl radical in antioxidant assays are not unheard of – Lichtenthäler et. al.(*3*), using the TOSC assay, also found results difficult to obtain using this radical species.

The SIFT-MS-TOSC assay is not as amenable to high-throughput analysis of samples as assays based on optical measurements, as each optical measurement may take less time than a SIFT-MS measurement and a lower volume of reaction mixture is necessary. However, optical-based antioxidant assays do not have the same flexibility of sample matrix as the SIFT-MS-TOSC assay, as they must all be conducted in aqueous solution. The ability to analyse aqueous and emulsified samples makes the SIFT-MS-TOSC assay a valuable addition to the list of antioxidant assays currently in use worldwide.

The SIFT-MS-TOSC assay must be conducted in a laboratory setting. Most laboratories would already contain all necessary equipment for this assay except of course, for a SIFT-MS instrument. These instruments are too expensive at the moment to justify a purchase solely for the SIFT-MS-TOSC assay, however if a laboratory has a SIFT-MS instrument available for other applications, performing the SIFT-MS-TOSC assay would be simple.

It is hoped that the current debate and uncertainty regarding the proper interpretation of antioxidant assay results and the search for standardised methods for antioxidant analysis will involve consideration of all methods, with final decisions based more on scientific rigour than economic or time constraints. There are currently several widespread antioxidant assays; it would be a shame if these were the only tests considered for a standardised method. The SIFT-MS-TOSC assay and other new-comers to the expanding field of antioxidant analysis may have some valuable contributions to make.

## 10.3 References

(1) Gomez-Alonso, S.; Salvador, M.; Fregapane, G. Evolution of the Oxidation Process in Olive Oil Triacylglycerol under Accelerated Storage Conditions (40-60°C). *J. Amer. Oil Chem. Soc.* **2004**, *81*, 177-184.

(2) Angerosa, F. Influence of Volatile Compounds on Virgin Olive Oil Quality Evaluated by Analytical Approaches and Sensor Panels. *Eur. J. Lipid Sci. Technol.* **2002**, *104*, 639-660.

(3) Lichtenthäler, R.; Marx, F.; Kind, O. M. Determination of Antioxidative Capacities Using an Enhanced Total Oxidant Scavenging Capacity (TOSC) Assay. *Eur. Food Res. Technol.* **2003**, *216*, 166-173.

# Appendix A

## Multivariate Statistical Techniques

### A.1 Introduction

In situations when many variables are being measured for many samples, data analysis is often difficult and time-consuming. Multivariate statistical techniques such as those described in this appendix offer significant advantages over simple visual inspection of data. They identify relationships between variables which can be used to summarise data or predict values for selected attributes in new samples. Several powerful techniques are available, each with their own benefits depending on the data collected and the desired results.

Two multivariate statistical techniques were employed during the course of this research. They are Principal Component Analysis (PCA) and Partial Least-Squares (PLS) regression. These are both unsupervised techniques: PCA is an exploratory technique, while PLS regression is confirmatory. Unsupervised techniques require only data as input, no operator manipulation is necessary. The data with which all multivariate statistical techniques begin are presented in the form of matrices. In the present research (as shown in figure A.1), all matrices are arranged with columns representing variables (e.g. VOCs measured or sensory attributes assessed) and rows representing cases (e.g. the value of each variable for each oil sample analysed).

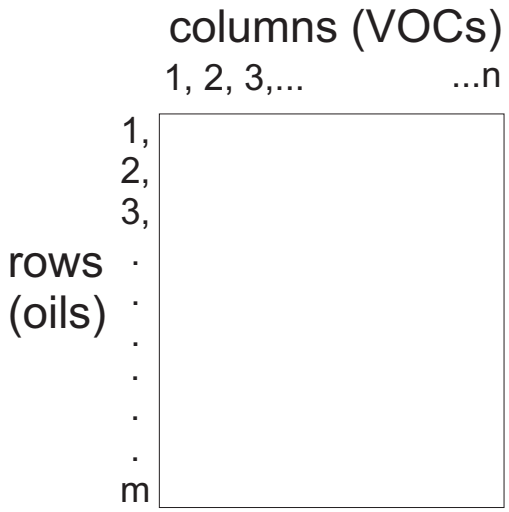



**Figure A.1.** An example of the arrangement of data into matrices in the present research. The total number of oils (rows) is denoted m, while the total number of VOCs measured (columns) is n.

Exploratory techniques are those which highlight relationships between variables or objects, while confirmatory techniques may be used to provide predictions. Similar processes are performed by the methods of PCA and PLS regression. The differences between the methods lie in the types of relationships highlighted (determined by the end goal of each method) and the point at which each algorithm ends. In the discussion that follows, an account will be given of the techniques of PCA and PLS regression. Both are commonly applied by practitioners of multivariate statistical methods.

### A.1.1 Summary of Matrices Used

In this appendix, seventeen matrices and thirteen vectors are introduced. They are catalogued here, with a brief explanation of the function of each. The use of all matrices and meaning of all terms are explained later in this appendix. This

summary is meant as a point of reference accompanying the later sections of this appendix. Matrices and vectors in this text are indicated by bold type: matrices by capital letters and vectors by lower case letters.

**X** is the matrix of original data shown in figure A.1, with rows organised by oil and columns organised by VOC.

**S** is the covariance matrix of the data in **X**. **S** shows the relationships between the VOCs in **X**, with both rows and columns organised by VOC.

**G** is the matrix of deviate scores of **X**. **G** is very similar to **X**, the only difference being that each column has had its mean value subtracted from each of its elements so that its mean is now zero.

**Z** is the standardised deviate score matrix of **X**. Like **G**, the mean value of each column is zero, however each column in the **Z** matrix has also been divided by the standard deviation of its elements so that they now have a standard deviation of one.

**R** is the correlation matrix of the data in **X**. **R** is very similar to **S**, only it shows correlations instead of covariances between VOCs.

**C** is the principal component loadings matrix. **C** gives important information about how the principal components relate to the VOCs.

**F** is the principal component scores matrix. **F** is similar to the **X** matrix, with rows organised by oil, only with the columns organised now by principal component as opposed to VOC.

**A** is the pattern matrix. **A** is a matrix of scaled eigenvectors, showing the directions of the principal component axes in relation to the VOC axes.

**U** is the matrix of eigenvectors. Each column of **U** is scaled by its respective eigenvalue to produce the **A** matrix.

**Λ** is the matrix of eigenvalues. Each diagonal element is the eigenvalue corresponding to the eigenvector of the principal component concerned, while each off-diagonal element is zero.

**D** is the diagonal standard deviation matrix. Each diagonal element in **D** is the standard deviation of the corresponding VOC measurements, while each off-diagonal element is zero.

**E** is a matrix of random noise. It represents the variation in the original data which is unable to be modelled by multivariate statistical techniques.

**W** is the weight loading matrix. **W** contains correlations between the **Z** matrix and the **y** vector from which PLS regression is carried out.

**T** is the oil latent variable score matrix. **T** is similar to the **Z** matrix, with rows organised by oil, only with the columns organised now by latent variable as opposed to VOC.

**P** is the latent variable loadings matrix. **P** relates the latent variables to the VOCs and shows which VOCs are most important for prediction of the strength of oil fruitiness.

**H** is the leverage matrix. Each diagonal element corresponds to one oil and shows how far its VOC concentrations are from the mean of the data. The off-diagonal elements, while not zero, are not used here.

**Ž** is a new standardised matrix of VOC concentrations from which the strength of the fruitiness attribute is to be predicted for each oil. Each column of **Ž** has

had the mean value of its corresponding column in the original **X** matrix used for calibration subtracted from each element. Each element has also been divided by the standard deviation of the **X** elements from the corresponding column.

$\mathbf{y_{un}}$ is the vector of fruitiness intensity values given to each oil from sensory analysis.

**y** is the standardised vector of fruitiness intensity values for the oils. **y** is $\mathbf{y_{un}}$ where each element has had the mean of the values subtracted and has been divided by the standard deviation of the $\mathbf{y_{un}}$ values.

**b** is the PLS regression vector. **b** is premultiplied by **Ž** to predict the intensity of the fruitiness attribute for new olive oils.

**e** is a vector of random noise. It represents the variation in the fruitiness attribute values which is unable to be modelled by multivariate statistical techniques.

**w** is the weight loading vector. **w** contains correlations between the **Z** matrix and the **y** vector from which PLS regression is carried out.

**t** is the oil latent variable score vector. **t** is similar to the **Z** matrix, with rows organised by oil. One **t** vector is produced for each latent variable during PLS regression.

**p** is the oil latent variable loadings vector. **p** relates the latent variable concerned to the VOCs and shows which VOCs are most important for prediction of the intensity of oil fruitiness. One **p** vector is produced for each latent variable during PLS regression.

**v** is the fruitiness latent variable loadings vector. **v** relates the latent variable concerned to the intensity of the fruitiness attribute.

**h** is the leverage vector. It is made up of the diagonal elements of **H**, the leverage matrix. Each element corresponds to one oil and shows how far its VOC concentrations are from the mean of the data.

$\hat{\mathbf{y}}$ is the standardised vector of fruitiness intensities predicted for each oil from the $\check{\mathbf{Z}}$ matrix.

$\hat{\mathbf{y}}_{\mathbf{un}}$ is the scaled (or unstandardised) vector of fruitiness intensities obtained by multiplying each value of $\hat{\mathbf{y}}$ by the standard deviation of the values in the **y** vector, then adding the average value of the **y** vector to each element.

$\hat{\mathbf{y}}_{\mathbf{mc\ cal}}$ is the mean-centred vector of fruitiness intensities predicted from leave-one-out cross-validation for the original oils. The mean of the elements of the vector is zero, and the standard deviation is equal to that of the elements of the original **y** vector.

$\hat{\mathbf{y}}_{\mathbf{mc\ new}}$ is the mean-centred vector of fruitiness intensities predicted from PLS regression for a new set of oils. The mean of the elements of the vector is zero, and the standard deviation is equal to that of the elements of the original **y** vector.

## A.2 Principal Component Analysis (PCA)

PCA is used to reduce the number of variables required to describe a system. This is achieved by creating new variables(*1*). These new variables (principal components, also called ‘factors’ or ‘latent variables’) are based on the

original variables, yet do not necessarily represent physically relevant quantities. Reyment and Jöreskog(*1*) give a very good, concise description:

PCA "creates a minimum number of new variables, which are linear combinations of the original ones such that the new variables contain most or all of the information."

Suppose we measure the concentrations of two VOCs in the head spaces above ten oils and represent the data as a matrix **X** (table A.1). The two VOCs we have measured are the variables to be reduced. This example has been engineered so that the two VOCs will reduce to one principal component. If the data are plotted on a chart where the two variables are represented by mutually orthogonal axes (figure A.1.(a)), PCA determines a new set of mutually orthogonal axes which are oriented in such a way that the first principal component axis is parallel to the direction of maximum variation of the data (figure A.1.(b)). The 'direction of maximum variation' is in fact the direction which contains the highest concentration of data points due to the highest correlation between two or more variables. Therefore, each principal component describes one source of correlation in the data. The first principal component best describes the highest correlation between variables, while each successive principal component describes the highest correlation which is uncorrelated with any previous principal components (each new principal component axis must be orthogonal to all established principal component axes so that no two principal components describe the same correlation)(*2*). It is possible to generate principal component axes which are not orthogonal, however the analysis of results is more complicated and they may be

misleading, therefore only orthogonal principal components were used in this research and only they will be described here.

**Table A.1.** The data matrix **X**, concentrations of two VOCs above ten different olive oils.

| | VOC 1 | VOC 2 |
|---|---|---|
| Oil 1 | 10.1 | 5450 |
| Oil 2 | 21 | 8540 |
| Oil 3 | 11.8 | 6650 |
| Oil 4 | 27.6 | 10500 |
| Oil 5 | 13.5 | 7030 |
| Oil 6 | 24.6 | 8570 |
| Oil 7 | 21 | 8420 |
| Oil 8 | 14.9 | 6410 |
| Oil 9 | 4.68 | 4920 |
| Oil 10 | 19.6 | 7900 |

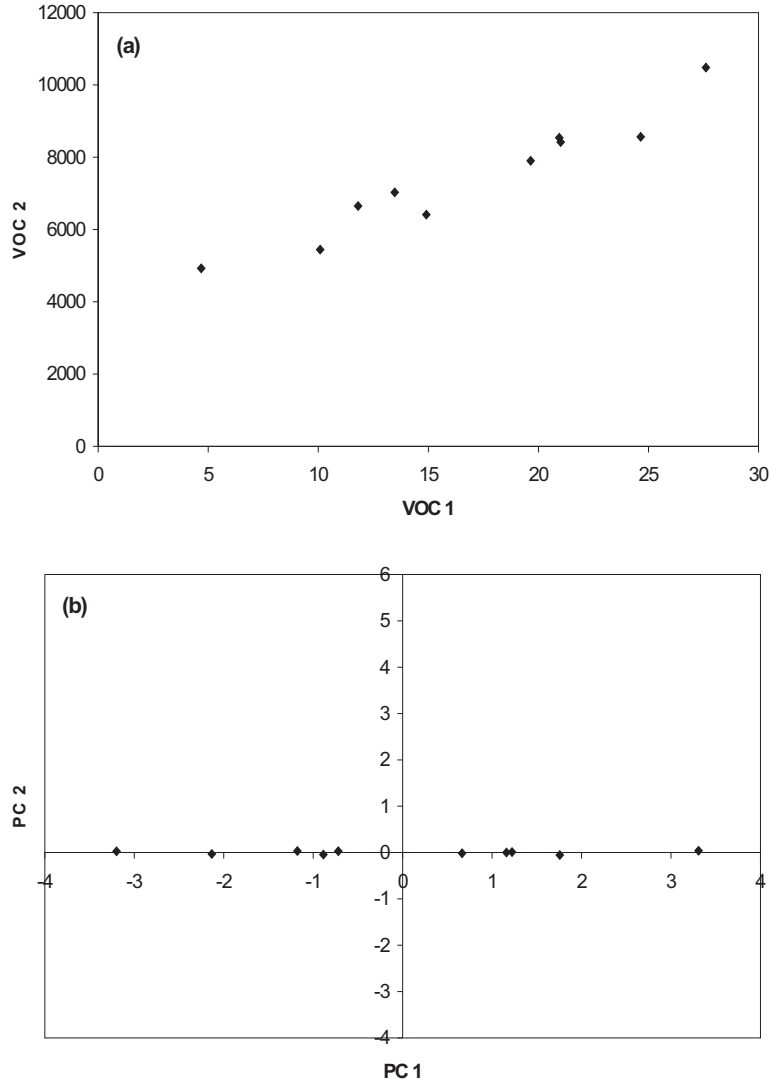



**Figure A.1.** The data from table 1 are shown **(a)** plotted by the VOCs measured (on the variable axes) and **(b)** plotted on the principal component axes constructed using PCA. Note that all data are now very close to the horizontal axis.

The core of PCA is the production of a covariance matrix (or correlation matrix), and the derivation of the eigenvectors and eigenvalues of that matrix. The eigenvectors show how to construct the principal components and the eigenvalues

show how important each one is. A covariance matrix is a matrix in which each entry (called an element) is a measure of the relationship between two variables, called the covariance. Correlation, a more familiar term to the uninitiated, is the covariance divided by the product of the standard deviations of the two variables being compared. The covariance matrix **S** is constructed from the matrix of deviate scores **G** (equation A.3). The term 'score' is used to signify a value that has no direct physical relevance, i.e. the values are no longer concentrations. The deviate score matrix **G** is constructed simply from the original data matrix **X** by subtracting the average concentration value of each VOC from the concentration measured for each oil (equation A.1(*1*)).

$$g_{ij} = x_{ij} - \mu_j \qquad \text{(A.1)}$$

In equation A.1, $g_{ij}$ is the element of **G** in row i (representing the objects or oils measured) and column j (representing the variables or VOCs), $x_{ij}$ is the corresponding element of **X**, and $\mu_j$ is the average value of the elements in column j of **X** (the average value measured for the VOC represented by row j). Each column of the **G** matrix has mean zero, each element is the deviation of each measurement for each oil from the mean value measured for each VOC (hence the name deviate score matrix). The deviate scores of the data from table A.1 are shown in figure A.2.

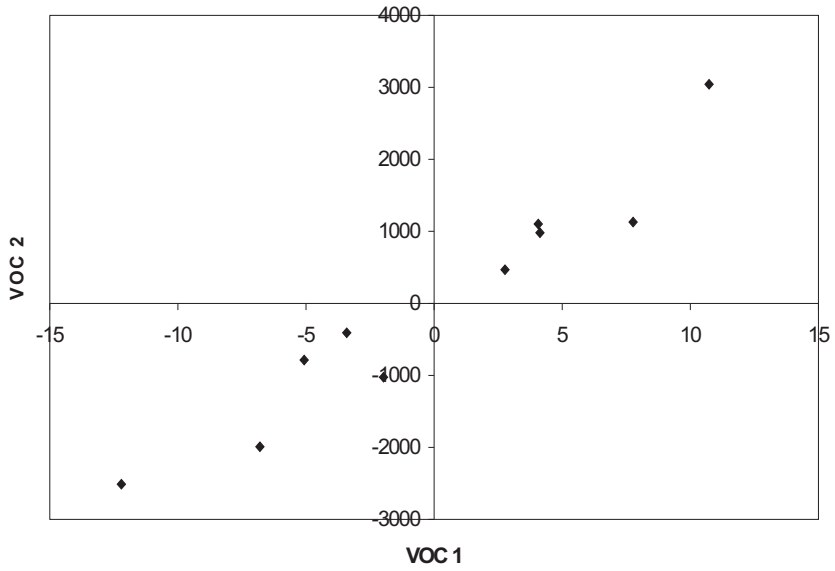



**Figure A.2.** The data from table A.1 and figure A.1.(a) mean-centred, as found in the deviate score matrix **G**.

If the ranges of all variables are not comparable (if some variables have much larger deviations than others), yet each provides comparable information, a **Z** (standardised deviate score) matrix may be constructed in place of a **G** matrix. This is a choice which must be made by inspection of the data, and an understanding of the sampling method is necessary(*3*). A **G** matrix will produce a model which is weighted more heavily toward those variables which display higher variance (this option compares the covariances of the variables), whereas a **Z** matrix will produce a model which compares all variables equally (correlations are used). Using a **Z** matrix may be troublesome if certain variables only measure noise, as the variation in these variables will be inflated to match that of more important variables. Inflation of scores for unimportant variables may mean that important correlations

are more difficult to identify, or that sampling noise is mistaken for a correlation. A **G** matrix should be used if all variables measured are rated on a similar scale, such as standard olive oil tasting data, which are evaluated on a scale from zero to ten. In this situation one variable may display little variance, representing a true lack of variance in the data. Inflating the importance of such variables by use of a **Z** matrix may produce a model which does not describe the data well. As described, **G** and **Z** matrices can give very different results. It has been suggested(*3*) that a **G** matrix be used unless there is a reason for the use of a **Z** matrix, although, time permitting, it is interesting to use both and compare the results.

The values for each VOC in our **Z** matrix have a mean of zero and a standard deviation of one, as is shown in equation A.2(*1*).

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \qquad \text{(A.2)}$$

Where $z_{ij}$ is the element of **Z** in row i and column j and $\sigma_j$ is the standard deviation of the elements of column j. The standardised deviate scores of the data from table A.1 are shown in figure A.3.

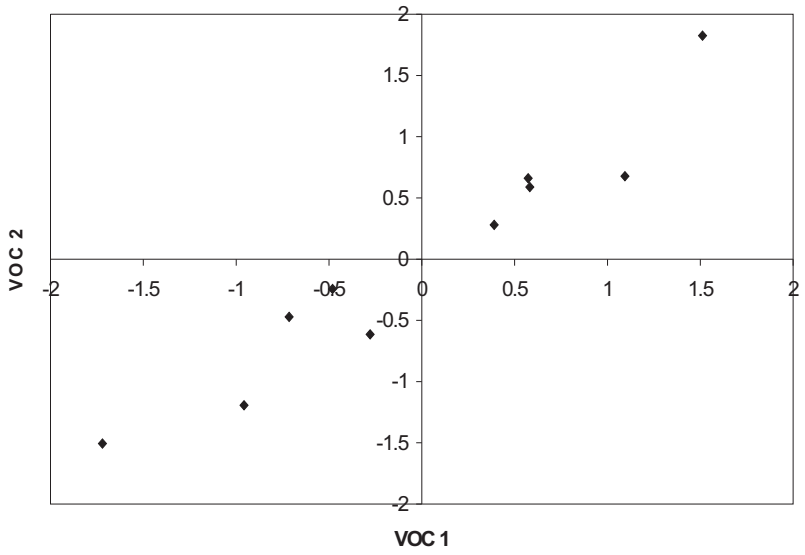



**Figure A.3.** Data from table A.1 and figure A.1.(a) standardised, as found in the standardised deviate score matrix **Z**. This figure is very similar to figure A.2, only the axis values have changed.

The covariance matrix **S** is constructed as shown in equation A.3(*1*).

$$\mathbf{S} = \frac{\mathbf{G}^{\mathbf{T}}\mathbf{G}}{\mathrm{m} - 1} \qquad \text{(A.3)}$$

Where m is the number of oils measured and $\mathbf{G}^{\mathbf{T}}$ is the matrix transpose of **G** (transposing flips a matrix about its diagonal, transforming a 10 x 2 matrix into a 2 x 10 matrix).

**Table A.2.** Correlation (**R**) and covariance (**S**) matrices for the VOCs of the data shown in table A.1. The trace of **R** (the sum of its diagonal elements) is 2, while the trace of **S** is 2.78 x $10^6$.

| **R** | | | **S** | |
|---|---|---|---|---|
| 1 | 0.966 | | 2780000 | 11400 |
| 0.966 | 1 | | 11400 | 50.3 |

If the **G** matrix is replaced by a **Z** matrix, the covariance matrix – although calculated in the same way – becomes a correlation matrix, **R**. The covariance matrix is a square symmetric matrix ($\mathbf{S} = \mathbf{S}^T$) which has diagonal elements equal to the standard deviations of the corresponding variables squared (their variances), and off-diagonal elements equal to the covariance between the corresponding variables. The correlation matrix is the same although due to standardisation, each of its diagonal elements is equal to one, and off-diagonal elements are between 1 and -1, indicating positive and negative correlations respectively between variables.

Now that the **S** (or **R**, if standardised data are used) matrix has been generated, its eigenvectors and eigenvalues may be found. An eigenvector is a vector which is aligned with the direction of greatest variation in the data set from which it was constructed. The standard mathematical definition of an eigenvector **u** is any of a set of non-zero vectors which satisfies equation A.4(*1*):

$$\mathbf{Ru} = \lambda\mathbf{u} \qquad \text{(A.4)}$$

Where **R** is a square matrix and $\lambda$ is a specific scalar, called the eigenvalue of the eigenvector. Put another way, postmultiplying a matrix by one of its

eigenvectors gives a vector which is proportional to the eigenvector (i.e. it does not change its direction). It often does, however, change its length. The factor by which its length is multiplied is called the eigenvalue of the eigenvector. Most standard mathematics computer programs include a function for the calculation of eigenvectors and eigenvalues of square matrices – the process is too involved to be described here.

A set of eigenvectors equal in number to the set of VOCs measured is constructed. The direction of each eigenvector from the origin coincides with that which contains the highest density of data points. Each successive eigenvector is constrained to being orthogonal to all established eigenvectors. Orthogonality is achieved by subtracting all variation in the direction of an eigenvector axis from the data before generating the next eigenvector. The eigenvectors represent the directions of the principal component axes – they are orthogonal just as the VOC axes are orthogonal. The terms 'eigenvector' and 'principal component' are equivalent in this context. By convention, all eigenvectors have unit length, as only the direction of an eigenvector is important. As mentioned above, each eigenvector has a corresponding eigenvalue. The size of this eigenvalue is a measure of how much variation is represented by the eigenvector (how much of the variation in the data set is projected along the direction of the eigenvector). The sum of all eigenvalues is equal to the sum of the squared elements of the trace (the diagonal elements) of the **S** matrix(*1*). Therefore, if an **R** matrix is constructed instead (where all diagonal elements are one), the sum of all eigenvalues obtained is the same as the number of VOCs measured.

The aim of PCA is to simplify data by allowing it to be expressed using fewer variables than it was originally. The principal components are to be our new variables, however we currently have the same number of principal components as we had VOCs to begin with. It may seem that our efforts to date have not improved the situation. Whether this is true or not comes down to the underlying nature of the original data. Considering figure A.1.(a), it can be seen that there is only one underlying factor in the data being measured, and that unknown factor is expressed as the correlation between VOC 1 and VOC 2. The unknown factor may be any aspect of the oil, for example its age, degree of exposure to micro-organisms, etc. The important point is that the unknown factor is not able to be measured directly – we can only measure its effect on VOC concentrations. If it were possible to represent that single factor with a single variable and the two measured VOCs did not provide any additional information, the data could be accurately represented using only one variable instead of the original two.

The eigenvectors constructed from either the **R** or **S** matrix represent the underlying factors (principal components) of the data, and the eigenvalues provide information on how many we need to properly represent the data.

There are several different methods for deciding the appropriate number of principal components. One of the most widely used and most intuitive methods is the scree plot. The eigenvalues are plotted with even spacing on a chart from the highest value to the lowest value, and the slope of lines drawn between each pair of points is considered. The point after which the slope becomes close to zero is considered to represent the first insignificant eigenvalue, and its associated

principal component is discarded along with all subsequent principal components(*4*). Figure A.4 shows a scree plot for a different set of hypothetical data, as this type of analysis is much better demonstrated with more than two variables.

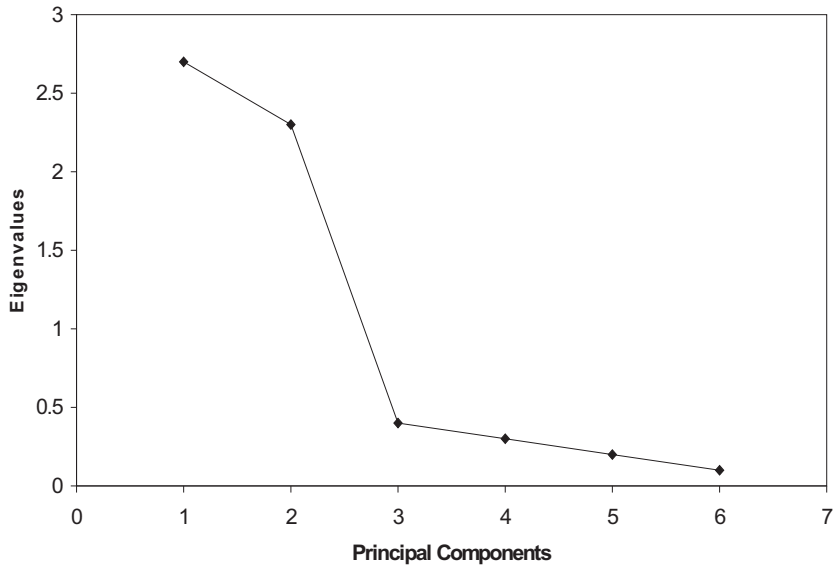



**Figure A.4.** A 'scree' plot of fabricated eigenvalues from a data set with six variables. The judgement of the point at which the slope becomes 'close to zero' is subjective, yet for these data the principal components corresponding to the four lowest eigenvalues may be safely discarded, leaving two.

Another method considers the proportion of the variation measured by each principal component. The proportion of each eigenvalue to the sum of all eigenvalues is the proportion of variation of the data found in the direction of the corresponding principal component. Starting at that which measures the most variation and proceeding in descending order, more principal components are added

to the model and the sum of eigenvalues for those principal components included is compared to the sum of all eigenvalues (the trace of the **R** or **S** matrix). When the sum of the eigenvalues included reaches a previously agreed level (say 95 or 99 % of the total), the principal components corresponding to the eigenvalues not included are regarded as noise and discarded(*1*). This method is illustrated in table A.3 using the same data as figure A.4.

**Table A.3.** Eigenvalues corresponding to each principal component and the variation measured. If 95 % of the variation is to be accounted for, four principal components are needed. For 99 %, all six must be included.

| Principal component | Eigenvalue | % variation measured |
|---|---|---|
| 1 | 2.7 | 45 |
| 2 | 2.3 | 83.3 |
| 3 | 0.4 | 90 |
| 4 | 0.3 | 95 |
| 5 | 0.2 | 98.3 |
| 6 | 0.1 | 100 |

The fact that several methods exist for principal component selection indicates the difficulty in selecting a point at which to restrict the model. This is still very subjective, and in practice a combination of these two methods along with knowledge of the sample will produce the most satisfactory model. If the number of significant principal components is two or three, the data may be plotted to produce an intuitive visual representation.

Considering the data from table A.1, whether we have standardised or not, only one principal component is significant. The relevant details are in table A.4.

**Table A.4.** Eigenvalues for each principal component of the data from table A.1 based on correlations (**R** matrix, labelled 'corr') and covariances (**S** matrix, labelled 'cov'). When correlations are considered, the second principal component measures less than 2 % of the variation and may be discarded. In the covariance case the second principal component is completely insignificant, hence should be discarded here also.

| Corr or cov? | Principal component | Eigenvalue | % variation measured |
|---|---|---|---|
| corr | 1 | 1.97 | 98.3 |
| | 2 | 0.0345 | 1.73 |
| cov | 1 | 2780000 | 99.9999 |
| | 2 | 3.42 | 0.000123 |

At the moment there is no way to determine which VOCs are most closely correlated with the principal components. It would also be interesting to know how each individual oil rates on this new variable. Both of these objectives may be met by the construction of new matrices. These are the **C** (principal component loadings) matrix and the **F** (principal component scores) matrix.

The **C** matrix contains correlations between each principal component axis and each original variable (VOC) axis. A value close to 1 or -1 for a particular variable indicates that the principal component is closely correlated with that particular VOC, or the VOC has a high 'loading' on the principal component. The signs of the principal components are arbitrary, so there is no real difference between negative and positive loadings, except that negative and positive loadings on the same principal component are different from each other. There is no set value to identify the VOCs with significant correlations, and all recommendations in the literature depend on the number of oils analysed. As the number of analysed

oils increases, the minimum value necessary for a loading to be significant decreases. As with the selection of the number of significant principal components, judgement must be used, and the practical significance of the results considered.

The **F** matrix contains the objects of the original data set transformed onto the principal component axes. The **F** matrix is used to obtain a graphical representation of the results of PCA, and is a good way to determine whether the results (assuming they are statistically significant) have any practical significance(*1*).

There are several ways to calculate these matrices, one is given below(*1*):

$$\mathbf{A} = \mathbf{U_k}\sqrt{\mathbf{\Lambda_k}} \tag{A.5}$$

$$\mathbf{C} = \mathbf{D}^{-1}\mathbf{A} \tag{A.6}$$

$$\mathbf{F} = \mathbf{GA}\mathbf{\Lambda_k}^{-1} \tag{A.7}$$

Where **A** is the pattern matrix (the unscaled principal component loadings matrix), $\mathbf{U_k}$ is a matrix whose columns are the eigenvectors of the correlation (**R**, or the covariance **S**) matrix, $\mathbf{\Lambda_k}$ is a matrix with its diagonal elements equal to the eigenvalues of the **R** or **S** matrix and its off-diagonal elements equal to zero, **D** is a matrix with its diagonal elements equal to the standard deviations of the original variables and its off-diagonal elements equal to zero. The subscript ‘k’ for the eigenvector and eigenvalue matrices is the number of principal components

selected. The parts of the **U** and **Λ** matrices referring to the discarded principal components (all **U** columns greater than k and all rows and columns of **Λ** greater than k) are discarded. If standardised data have been used (**Z** matrix instead of **G** as initial input), **D** will be an identity matrix, meaning the **A** matrix need not be scaled to obtain the principal component loadings matrix **C** (because **C** = **A** already). The **C** and **F** matrices are shown in table A.5.

**Table A.5. C** (principal component loadings) and **F** (principal component scores) matrices based on correlations ('corr') and covariances ('cov'). Each is in fact a vector in this example, due to the discovery of only one significant principal component.

| Corr | | Cov | |
|---|---|---|---|
| **C** | **F** | **C** | **F** |
| 0.991 | -1.09 | 0.965 | -1.19 |
| 0.991 | 0.622 | 1 | 0.66 |
| | -0.598 | | -0.472 |
| | 1.68 | | 1.82 |
| | -0.366 | | -0.244 |
| | 0.894 | | 0.677 |
| | 0.59 | | 0.588 |
| | -0.451 | | -0.615 |
| | -1.63 | | -1.51 |
| | 0.338 | | 0.28 |

In order to represent the new data graphically, two variables would be ideal. We may use the discarded principal component for this purpose. The elements of the **F** matrix have been scaled (equation A.7). To appreciate how much variation exists along each axis, the scaling may be undone by multiplying the values in each column of the **F** matrix by their respective eigenvalues.

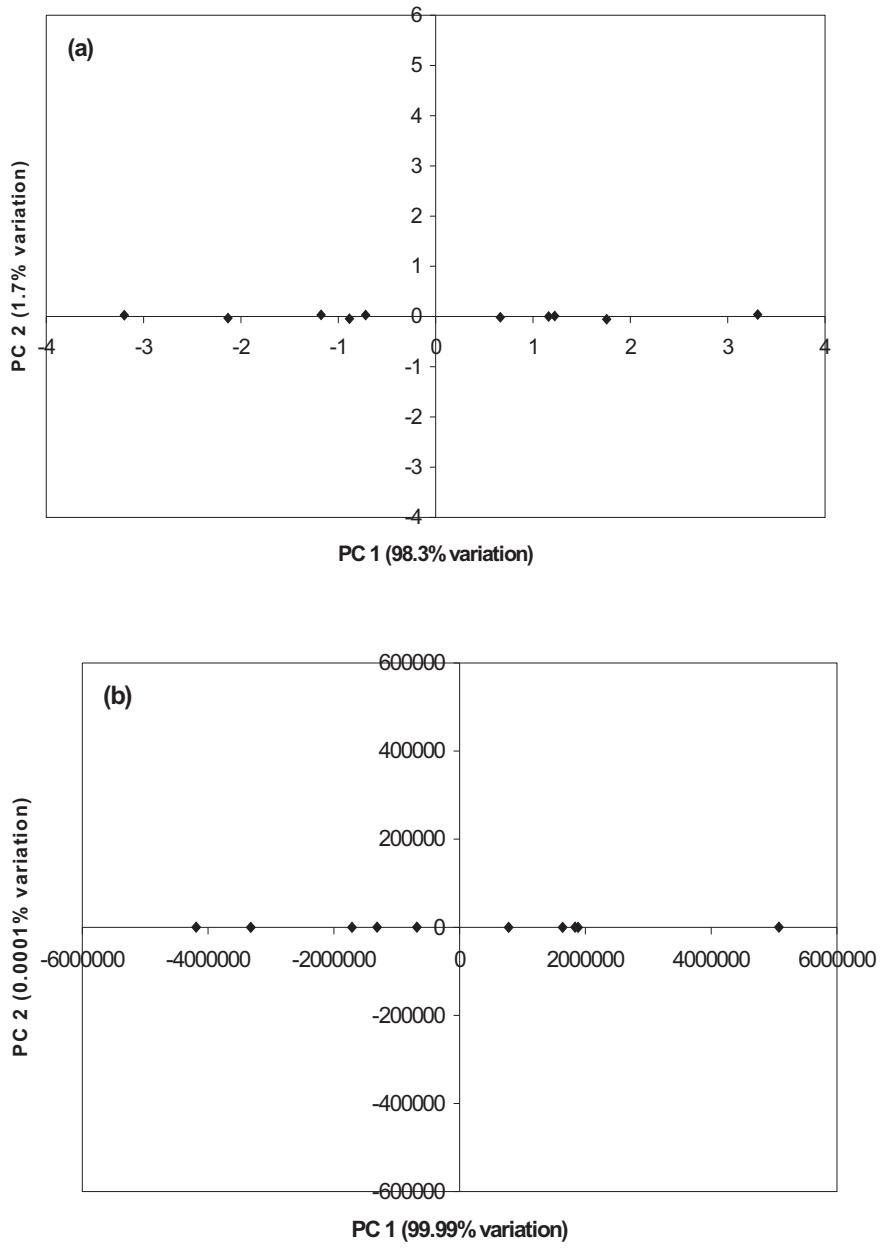



**Figure A.5.** Plots of rescaled principal component scores for data analysed by **(a)** correlations **(b)** covariances. The data are found almost exclusively along one axis. When correlations are considered (as in **(a)**), the significant principal component is aligned along a line of best fit for the points in figure A.3. The data shown here are projections of the original data onto this line. When covariances are considered (as

in **(b)**), the significant principal component is dominated by the highest concentration VOC (VOC 2), and the data shown here are projections of the original data almost solely onto that axis (although due to scaling the data have different absolute values). VOC 2 is not necessarily more important to the description of oil quality than VOC 1, so correlations are preferred over covariances in this example.

The **F** and **A** matrices in the covariance model now contain all the useful information that was part of **G**. Correspondingly, the **F** and **C** matrices in the correlation model contain all the useful information from **Z**. This can be demonstrated by the equation $\mathbf{G} = \mathbf{F}\mathbf{A}^{\mathbf{T}}$ (equally $\mathbf{Z} = \mathbf{F}\mathbf{C}^{\mathbf{T}}$). These equations are not completely correct, however, as these matrices are only approximations of the original data. The non-useful variation (noise) in the original data was removed from the **F** and **A** matrices (as they are least-squares approximations made from the original data), so a more correct equation would be $\mathbf{G} = \mathbf{F}\mathbf{A}^{\mathbf{T}} + \mathbf{E}$, where **E** represents the random noise contained in the principal components that were discarded earlier(*1*).

## A.3 Partial Least-Squares Regression

PLS regression is a confirmatory technique that shares some aspects with the exploratory technique of PCA. Both techniques can be used in situations where the number of measured samples m is approximately equal to, or even less than, the number of variables measured (n). Whereas PCA is used to simplify data, the purpose of PLS regression is prediction.

The same example from the PCA section of two VOCs measured for ten oils will be continued in this section. As the standardised data (contained in the **Z** matrix) provided the most acceptable results in PCA, only they will be used in PLS regression. PLS regression requires another oil attribute which the VOC concentrations will be used to predict. Fruitiness is a commonly evaluated sensory attribute of olive oil, assigned an intensity value between zero and ten. To ensure there is a significant relationship between fruitiness and VOC concentrations to discover, the degree of fruitiness of the oils will be made to correlate with the significant principal component identified by PCA. The VOC data are contained in the standardised **Z** matrix from the PCA section, and the vector containing the strength of the fruitiness attribute for each oil is $\mathbf{y_{un}}$ (table A.6). Once each member of $\mathbf{y_{un}}$ has had the mean subtracted and has been divided by the standard deviation ($\mathbf{y_{un}}$ is standardised), $\mathbf{y_{un}}$ becomes **y**.

**Table A.6.** Unscaled ($\mathbf{y_{un}}$) and standardised (**y**) fruitiness intensity values for ten hypothetical olive oils.

| | $\mathbf{y_{un}}$ | **y** |
|---|---|---|
| Oil 1 | 2.66 | -1.07 |
| Oil 2 | 7.03 | 0.88 |
| Oil 3 | 4.07 | -0.443 |
| Oil 4 | 8.61 | 1.58 |
| Oil 5 | 4.21 | -0.378 |
| Oil 6 | 7.38 | 1.03 |
| Oil 7 | 5.87 | 0.361 |
| Oil 8 | 3.43 | -0.728 |
| Oil 9 | 1.58 | -1.55 |
| Oil 10 | 5.77 | 0.314 |

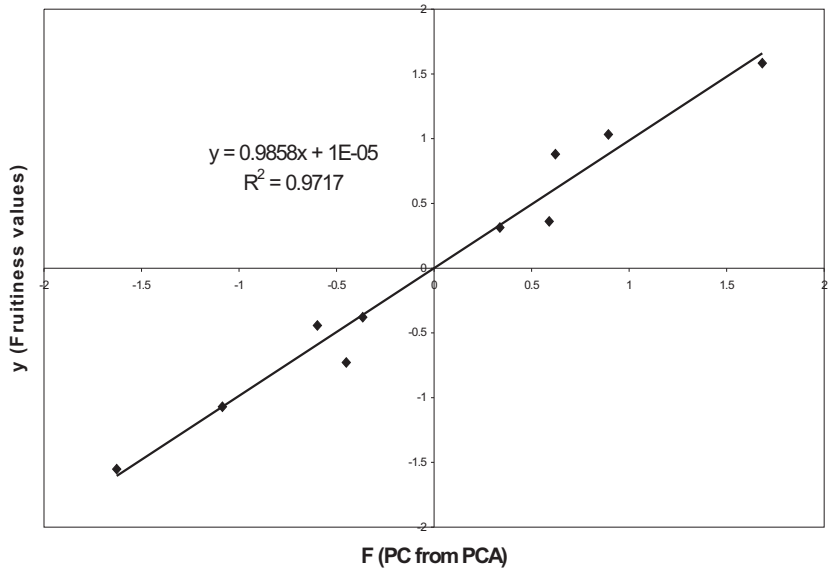



**Figure A.6.** The relationship between principal component loadings for the first component from PCA and standardised fruitiness intensity values for the present example involving ten olive oils.

**y** and **Z** are related to each other by a vector of coefficients **b** equation A.8:

$$\mathbf{y} = \mathbf{Zb} + \mathbf{e} \qquad \text{(A.8)}$$

Where **e** is a vector representing removed noise similar to the **E** matrix from the PCA section. This noise is not completely random as in the approximation involved in PCA. The vector **e** encompasses all variation in fruitiness values which cannot be approximated from VOC concentrations by using the coefficient matrix **b**. This includes variation caused by correlations with VOCs which were not measured, and correlations which are non-linear. The importance of the information contained in the **e** vector determines the degree of success of the investigation. If there is a lot of important information about **y** which is not included in the **Zb** term of equation A.8, satisfactory prediction will not be possible.

To derive **b** for the prediction of subsequent fruitiness values from VOC concentrations, equation A.9 is needed.

$$\mathbf{b} = \mathbf{Z}^{-1}\mathbf{y} \qquad \text{(A.9)}$$

Where $\mathbf{Z}^{-1}$ is the matrix inverse of **Z**. This is the matrix equivalent of a reciprocal, so that $\mathbf{Z}\mathbf{Z}^{-1} = \mathbf{I}$, where **I** is the identity matrix. In the identity matrix, all diagonal elements are one and all off-diagonal elements are zero.

The problem here is that to derive the regression vector **b**, the **Z** matrix of VOC concentrations must be mathematically inverted. **Z** has no unique inverse (a matrix with no unique inverse is termed 'singular') unless **Z** is square and no VOCs are correlated with each other(*5*). As we have seen from the previous section, **Z** is not square (its dimensions are 10 x 2) and the VOC concentrations are correlated.

So, we cannot use the simple inverse least-squares technique for these data. Luckily, there are several methods available for finding **b** that we can use.

In equation A.8, **e** may be discarded, as the information necessary to predict it is not available. Each side of the equation is premultiplied by $\mathbf{Z}^{\mathbf{T}}$ and the left and right terms are swapped for later simplicity, arriving at equation A.10.

$$\mathbf{Z}^{\mathbf{T}}\mathbf{Z}\mathbf{b} = \mathbf{Z}^{\mathbf{T}}\mathbf{y} \qquad \text{(A.10)}$$

Premultiplying by $(\mathbf{Z}^{\mathbf{T}}\mathbf{Z})^{-1}$ leaves an expression for **b** (equation A.11)(*6*):

$$\mathbf{b} = (\mathbf{Z}^{\mathbf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathbf{T}}\mathbf{y} \qquad \text{(A.11)}$$

Where $(\mathbf{Z}^{\mathbf{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathbf{T}}$ is the pseudo-inverse of **Z**, a least-squares approximation of $\mathbf{Z}^{-1}$.

Equation A.11 requires the inversion of a (unscaled) correlation matrix (cf. equation A.3). This will always be square, therefore one of the conditions for a successful matrix inversion has been satisfied. However, the VOCs are still correlated with each other. The most simple way to eliminate this problem is to eliminate the measurements for some VOCs from the original data matrix. A method which uses this approach is called Multiple Linear Regression (MLR)(*5*). All but a small number of variables which are uncorrelated with each other are discarded and the remaining variables used to construct the correlation matrix. It is

not difficult to imagine that the use of MLR can lead to loss of important descriptive information.

The condition which calls for uncorrelated variables can be satisfied by constructing eigenvectors. As described in the PCA section, eigenvectors are perfectly suited to this situation as they represent all important information and are uncorrelated. The eigenvectors can then be used to calculate **b**. Different eigenvectors from those calculated by PCA are needed for prediction in many cases, as those eigenvectors which best describe the relationships between VOCs may be irrelevant for predicting the intensity of the fruitiness attribute. Principal Component Regression (PCR) is a technique which attempts to use PCA eigenvectors for prediction. In some situations (such as the present example) it provides similar results to PLS regression, yet PLS regression is much more flexible than PCR. PLS regression eigenvectors are called latent variables here (to differentiate them from principal components derived from PCA), and are generated to provide the best possible prediction of the dependent variables (fruitiness).

The Nonlinear Iterative Partial Least-Squares (NIPALS) algorithm presented by Haaland and Thomas(*7*) was used as it is one of the two most commonly used PLS algorithms and involves the most straightforward mathematics. The NIPALS algorithm calculates the latent variables one at a time, subtracting the variation measured by each latent variable at the end of each cycle before creating the next latent variable. Instead of obtaining correlations between VOCs (using $\mathbf{Z}^T\mathbf{Z}/(N-1)$ or the **Z** correlation matrix) as in PCA, the latent variables

are constructed from the correlations between VOC concentrations and fruitiness values. Whether PCA or PLS regression are used, the two types of vectors required as output are loading and score vectors, which show the correlations between VOCs and latent variables and the value obtained by each oil on each latent variable respectively. For PLS regression an additional loading vector is necessary to show the correlation between the intensity of the fruitiness attribute and each latent variable.

The VOC concentrations and fruitiness values (**Z** and **y** respectively) are initially related to each other by **w**, which is given the vague title of the 'weight loading vector'. **w** is analogous to the correlation matrix in PCA from which the eigenvectors are calculated (equation A.12)(*7*). However, as there is only one variable with which the VOCs are being correlated, **w** is a vector instead of a matrix.

$$\mathbf{w} = \frac{\mathbf{Z}^T \mathbf{y}}{\mathbf{y}^T \mathbf{y}} \tag{A.12}$$

**w** is then normalised by dividing by its length (equation A.13).

$$\mathbf{w_{new}} = \frac{\mathbf{w_{old}}}{\sqrt{\mathbf{w_{old}}^T \mathbf{w_{old}}}} \tag{A.13}$$

The subscripts used here are simply to distinguish the two vectors which are both named **w**. $\mathbf{w_{old}}$ is the **w** vector from equation A.12, while $\mathbf{w_{new}}$ is the **w** vector used in all subsequent equations.

**t**, the oil score vector on the first latent variable, is calculated from **Z** and **w** in equation A.14(*7*). **t** here is analogous to the first column of the **F** matrix created in the PCA method, in that its elements may be plotted to show the arrangement of the oil samples with respect to the latent variables. Plotting these scores in PLS regression is not as valuable as in PCA, where principal component scores are among the most valuable output. The most important output of PLS regression is the b vector, as it allows prediction of fruitiness intensity for subsequent oils. To gain information on the relationship between the VOCs and the latent variable, the **p** vector should be consulted. The **p** vector is constructed in equation A.16.

$$\mathbf{t} = \frac{\mathbf{Zw}}{\mathbf{w}^T \mathbf{w}} \qquad \text{(A.14)}$$

v, the fruitiness loading on the first latent variable, is calculated from **t** and **y** (equation A.15)(*7*). Here v is a scalar (matrix dimensions of $\mathbf{t}^T\mathbf{y}$ = (1 x m) x (m x 1) = (1 x 1)). This is because only a single sensory attribute is being predicted. Using multivariate **Y** (e.g. predicting fruitiness and bitterness simultaneously) would produce a row vector **v**, with the same number of columns as there are sensory attributes. Using multivariate **Y** would also increase the complexity of the method and require some steps to be repeated. Predicting multivariate **Y** values often requires a compromise situation where some attributes are predicted less well

than they could be in order to provide the best overall prediction of the desired attributes. Each sensory attribute was predicted individually in this research, so the multivariate case will not be described here.

$$v = \frac{\mathbf{t}^T \mathbf{y}}{\mathbf{t}^T \mathbf{t}} \tag{A.15}$$

Once the latent variable scores have been calculated, **p**, the vector of VOC loadings on the first latent variable, is found (equation A.16)(*7*). The elements of **p** are the correlations between each VOC and the new latent variable. They are analogous to the elements of the **C** matrix in PCA (i.e. they show how large a contribution is made to a latent variable by each VOC).

$$\mathbf{p} = \frac{\mathbf{Z}^T \mathbf{t}}{\mathbf{t}^T \mathbf{t}} \tag{A.16}$$

With all the necessary information obtained for the first latent variable (table A.7), the variation in both the fruitiness intensities and the VOC concentrations which is measured by the latent variable (**t**v and $\mathbf{tp}^T$ respectively) is calculated. To ensure that subsequent latent variables are uncorrelated with previous latent variables, all variation measured by the latent variable is subtracted from the original data (equations A.17 and A.18)(*7*) before the vectors associated with the next latent variable are calculated.

**Table A.7.** The vectors generated from equations A.12-16 for the data used in the present example. **w** is the weight loading vector, **t** the latent variable scores vector, v the loading of fruitiness intensity on the latent variables and **p** the loading vector of VOC concentrations on the latent variables. **p** is the most useful with regard to simple inspection, as it contains the correlations of the VOCs with the latent variable (both VOCs show identical correlations here, meaning both VOCs are equally important for predicting fruitiness intensity), however **w**, **p** and v are all needed to construct the regression vector **b** (equation A.19).

| **w** | **t** | v | **p** |
|---|---|---|---|
| 0.704 | -1.52 | 0.703 | 0.707 |
| 0.71 | 0.873 | | 0.707 |
| | -0.838 | | |
| | 2.36 | | |
| | -0.512 | | |
| | 1.25 | | |
| | 0.828 | | |
| | -0.633 | | |
| | -2.28 | | |
| | 0.473 | | |

$$\mathbf{e}_y = \mathbf{y} - \mathbf{t}v \tag{A.17}$$

$$\mathbf{E}_Z = \mathbf{Z} - \mathbf{t}\mathbf{p}^T \tag{A.18}$$

$\mathbf{e}_y$ and $\mathbf{E}_Z$ are the vector of **y** and matrix of **Z** 'residuals' respectively. Each **w** and **p** vector and v value generated is stored for later use. The algorithm is then begun anew, with $\mathbf{e}_y$ and $\mathbf{E}_Z$ substituted for **y** and **Z** respectively. As each latent variable is calculated, $\mathbf{e}_y$ and $\mathbf{E}_Z$ contain less important information. Once the

correct number of latent variables has been calculated, $\mathbf{e}_y$ is equal to **e** from equation A.8 for these data.

To enable prediction of fruitiness intensity from VOC concentrations for new data ($\hat{\mathbf{y}}$ and $\check{\mathbf{Z}}$ respectively), the regression vector (**b**) must be generated. Haaland and Thomas(*7*) demonstrate the calculation of the **b** vector in PLS regression.

The **w** and **p** column vectors and the v values corresponding to the latent variables generated are all separately concatenated (pasted together) to give **W** and **P** matrices and a **v** vector respectively, each with the same number of columns as there are latent variables. The regression vector **b** is then calculated (equation A.19, cf. equation A.11):

$$\mathbf{b} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{v}^T \qquad \text{(A.19)}$$

$\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}$ is now the pseudo-inverse of **Z** as described by the latent variables and $\mathbf{v}^T$ (the **y** loading vector) corresponds to an approximation of **y**. The regression vector for the present example is shown in table A.8.

**Table A.8.** The regression vector **b** calculated for predicting fruitiness intensity. Its rows are the values by which VOC concentrations are multiplied to achieve the best prediction of fruitiness intensity values.

| | **b** |
|---|---|
| VOC 1 | 0.4951 |
| VOC 2 | 0.4993 |

Once the regression vector is obtained, $\hat{\mathbf{y}}$ is simply calculated from $\check{\mathbf{Z}}$ (equation A.20, cf. equation A.8):

$$\hat{\mathbf{y}} = \check{\mathbf{Z}}\mathbf{b} \qquad \text{(A.20)}$$

It must be remembered that since $\mathbf{Z}$ is standardised, $\check{\mathbf{Z}}$ also must be standardised to the same scale (for each VOC recorded in $\mathbf{X}$ the mean concentration from the original calibration data must be subtracted from the concentration for all oils and the resultant values divided by the standard deviation from the original calibration data) and the $\hat{\mathbf{y}}$ values calculated will be on the same scale as the $\mathbf{y}$ data. Therefore, to obtain physically meaningful values, each member of $\hat{\mathbf{y}}$ should be multiplied by the standard deviation and added to the mean of the original $\mathbf{y_{un}}$ data used for calibration.

PLS regression is also similar to PCA in other ways: it creates the same number of latent variables as there are VOCs in the original data, and deciding the correct number to keep is not always an easy task. Not all of the variation in the fruitiness values will be accounted for in PLS regression, meaning that the criteria used in PCA are not suitable here. The solution in PLS is more practical.

As the goal is the best possible prediction of the $\hat{\mathbf{y}}$ values, the best number of latent variables to include is that which provides the best prediction. A common method of evaluating this involves the use of a new ‘test’ set of data, separate from the data already used, for which the fruitiness values are known. The known values are compared with values predicted using different numbers of latent variables.

Leave-one-out cross-validation is a common method for the generation of a test set(*7*). This method allows the best possible use of the acquired data, as only the calibration set is needed. All data concerning one oil are removed from the original data and the procedure (equations A.12-18, performed the same number of times as there are VOCs) is carried out for the remaining data. The fruitiness value for the removed oil is predicted (equations A.19 and A.20) and the 'residual' (the difference between the values) is recorded. The data for that oil are then re-inserted, so that a set of m-1 oils is always used. This is repeated for each oil, resulting in the generation of the same number of regression (**b**) vectors as there are oils (m, which in this example is 10) and a vector of residuals with that same number of elements. The predicted fruitiness intensity values and residuals from leave-one-out cross-validation are shown in table A.9.

**Table A.9.** The vectors $\mathbf{y_{un}}$ and $\mathbf{\hat{y}_{un}}$ (measured and predicted, respectively) are fruitiness intensity vectors. The residual values from leave-one-out cross-validation may be used to evaluate the predictions.

| $\mathbf{y_{un}}$ | $\mathbf{\hat{y}_{un}}$ | residuals ($\mathbf{y_{un}}$ -$\mathbf{\hat{y}_{un}}$ ) |
|---|---|---|
| 2.66 | 2.66 | 0.006 |
| 7.03 | 6.34 | 0.7 |
| 4.07 | 3.68 | 0.39 |
| 8.61 | 8.94 | -0.33 |
| 4.21 | 4.26 | -0.045 |
| 7.38 | 6.97 | 0.41 |
| 5.87 | 6.45 | -0.57 |
| 3.43 | 4.15 | -0.72 |
| 1.58 | 1.38 | 0.2 |
| 5.77 | 5.81 | -0.047 |

Important information may be obtained from the vector of residuals generated by leave-one-out cross-validation. An abnormally large residual often signifies an outlier. Outliers can often be identified by way of a plot of measured versus predicted fruitiness intensity values, although there are also more formal tests(*8*). An outlier is an oil that displays a value for an attribute which is not able to be adequately predicted, as the necessary information is not available in the **Z** matrix of VOC concentrations. The value may be correlated with VOCs which were not measured, or perhaps correlated with species which are non-volatile. Outliers can be dangerous in PLS regression, as several latent variables may have been altered considerably to account for a few oils which are not representative of the population. If not removed, the outliers could seriously inhibit the ability to predict fruitiness values for new oils. If there is large variation in **b** vectors and residuals found during leave-one-out cross-validation, the possibility that the data are not well suited to PLS regression must be considered and alternative methods explored.

There are several measures which are used to evaluate the results of leave-one-out cross-validation to determine the most appropriate number of latent variables for prediction of fruitiness intensity. Höskuldsson(*9*) discussed different measures of latent variable significance and noted that none will provide optimum results for all data sets. Two measures were used for evaluation in this research, and each will now be described.

The Root Mean Square Error of Cross-Validation (RMSECV) is a measure of the magnitude of prediction residuals(*10*). It is the square root of the of the

Prediction Error Sum of Squares (PRESS) divided by the number of oils (equation A.21). The PRESS is the sum of the squared deviations of the tested points (y) from the predicted points ($\hat{y}$) for all m oils as each is left out, and is shown as the numerator in equation A.21:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{m}(y - \hat{y})^2}{m}} \qquad \text{(A.21)}$$

The **y** and $\hat{\mathbf{y}}$ vectors should be rescaled to give physically relevant values for the fruitiness intensity ($\mathbf{y_{un}}$ and $\hat{\mathbf{y}}_{\mathbf{un}}$) before the RMSECV is calculated. If the RMSECV is calculated using standardised data, not only are the absolute values different, they are different relative to each other. This difference may cause a sub-optimum number of latent variables to be recommended, which in turn will not provide the best possible predictions. The RMSECV values calculated for the present example are shown in table A.10.

**Table A.10.** RMSECV and $R^2$ values for the determination of the optimum number of latent variables to include for prediction of fruitiness intensity values. Lower RMSECV and higher $R^2$ values denote better prediction. Both measures here recommend the use of one latent variable as opposed to two.

| Number of latent variables | RMSECV | $R^2$ |
|---|---|---|
| 1 | 0.425 | 0.961 |
| 2 | 0.515 | 0.943 |

It is desirable to include as many latent variables as are necessary to give the lowest RMSECV, as this will provide the best predictions. There is always random variation (noise) in the data – this is exposed by the leave-one-out method. In most cases the last few latent variables will measure this noise, causing the RMSECV to rise. These latent variables should be discarded. After this there are commonly latent variables left which provide only small improvements and are unnecessary. Including extra latent variables adds time to calculations and makes the loadings matrix (**P**) more difficult to interpret. To be worth the extra effort, each successive latent variable should provide a significant improvement (a drop of at least 2% in the RMSECV is recommended by the Matlab PLS Toolbox Reference Manual(*10*)) in order to be included. Following this advice for the data presented in table A.10, as expected, one latent variable is recommended.

The second evaluation procedure used was consideration of the $R^2$ value. Plotting the measured values against the predicted values of fruitiness intensity is a straightforward method for evaluating the prediction power. As the ultimate goal of PLS regression (and indeed all regression) is to predict values of the desired variable with the greatest accuracy and precision possible, this method appears very

appropriate for evaluation. The predictive power obtained by including different numbers of latent variables may be compared by calculating the $R^2$ values when the predicted values are plotted against the measured values (table A.10, also figure A.7). This method is not as simple to evaluate as the RMSECV approach, as a significant $R^2$ value does not necessarily mean that the two values being compared are the same in all instances. The slope and intercept of this plot must also be considered to ensure that they are close to one and zero respectively. The $R^2$ value is still useful, however, as it is a very familiar measure of predictive power.

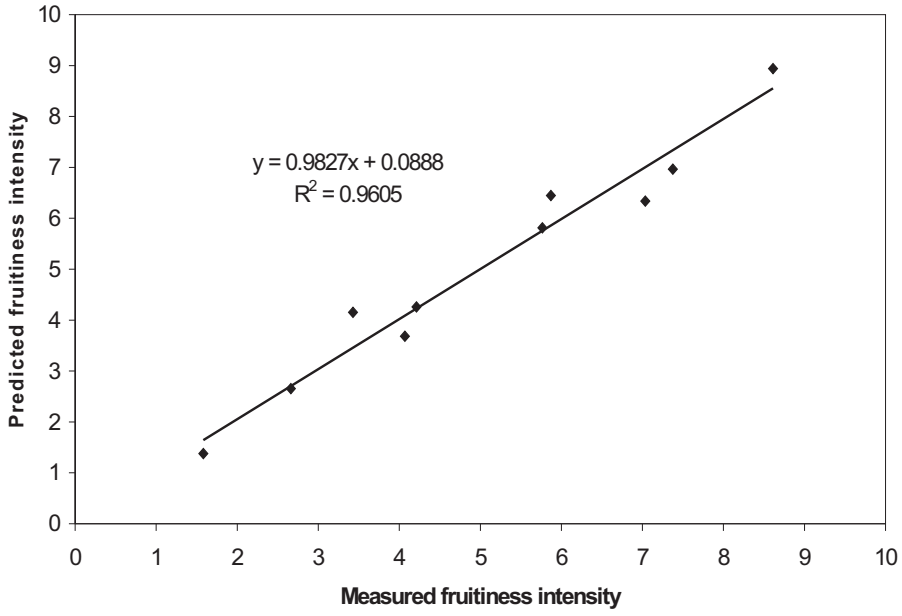



**Figure A.7.** The relationship between measured and predicted fruitiness intensity using a regression vector (**b**) derived from the first latent variable. The $R^2$ value for prediction is 0.96.

Both of the evaluation methods described above were used to evaluate PLS regression predictions, following the caution from Höskuldsson(*9*) that different

methods may recommend different numbers of latent variables. Different methods are best in different situations, therefore several should be used and the resultant predictions compared.

Caution is necessary, however, as over-fitting may be observed for some data sets. Over-fitting results in a very good fit for the original data used, yet a poor fit for the prediction of new samples(*11*). This is the result of poor sampling, where the calibration set is not a good representation of the population, it is not a fault of the technique used. Over-fitting may be caused by sampling bias such as selecting only abnormal oils, or an insufficient sample size, where adequate representation of the population is very difficult. Over-fitting may also be observed if the data are inadvertently taken from a different population from that to which the results are to be applied. In this case, re-sampling is the best solution. Over-fitting due to biased samples and insufficient sample size caused concern in the present research.

Prediction errors associated PLS regression may be shown by constructing a prediction interval for each point. A prediction interval is a range constructed around a predicted value within which there would be a given probability of finding the corresponding measured value if it were measured. Prediction intervals may be constructed for the results of leave-one-out cross-validation to give a convenient indication of the size of the intervals before unknown samples are analysed. A method for constructing prediction intervals was given by Faber et. al.(*12*), with supplementary information from Gemperline(*8*). The necessary equation for each predicted fruitiness intensity value is shown below (equation A.22)(*12*), with

prediction intervals calculated for the predicted fruitiness intensities of the leave-one-out cross-validation samples listed in table A.11 and plotted in figure A.8.

**Table A.11.** Prediction intervals at 95 % confidence for cross-validation fruitiness intensity values calculated via equations A.22 and A.23. The appropriate value from Student's t distribution is 1.86 ($\alpha = 0.05$, $m = 10$) and the standard deviation was calculated from the residuals in table A.9.

| Predicted value | Lower limit | Upper limit |
|---|---|---|
| 2.66 | 1.77 | 3.54 |
| 6.34 | 5.49 | 7.19 |
| 3.68 | 2.83 | 4.53 |
| 8.94 | 7.98 | 9.9 |
| 4.26 | 3.42 | 5.1 |
| 6.97 | 6.1 | 7.83 |
| 6.45 | 5.59 | 7.3 |
| 4.15 | 3.31 | 4.99 |
| 1.38 | 0.429 | 2.33 |
| 5.81 | 4.97 | 6.65 |

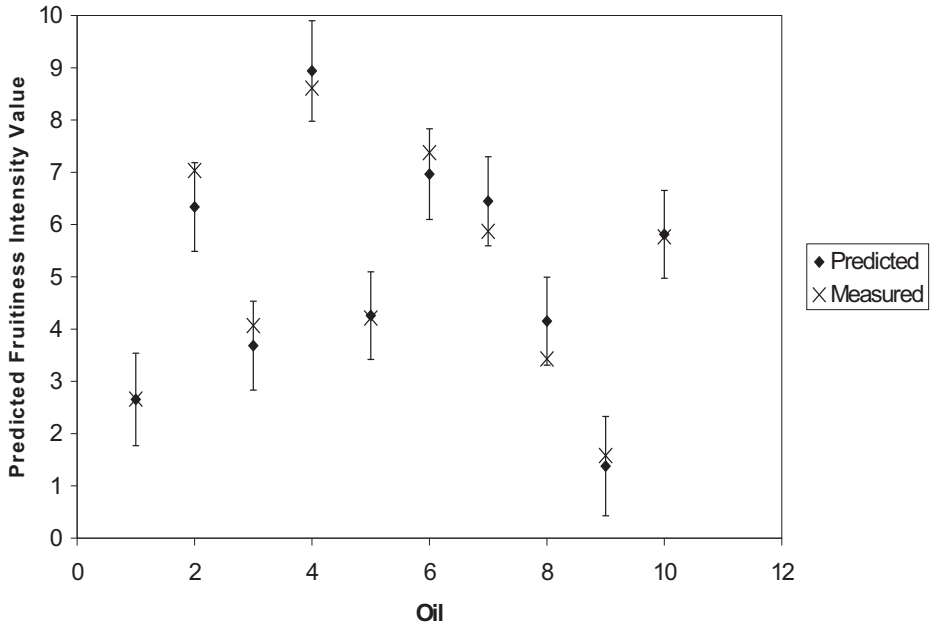



**Figure A.8.** Predicted and 'measured' values of fruitiness intensity plotted with prediction intervals at 95 % confidence from table A.11.

$$PI = \hat{y} \pm t_{m-2,1-\frac{\alpha}{2}}\, s\sqrt{1+h} \qquad (A.22)$$

Where PI represents the values at the high and low ends of the prediction interval, $\hat{y}$ is the predicted fruitiness intensity value, $t_{m-2,1-\alpha/2}$ is the critical value from Student's t distribution with m-2 degrees of freedom at the $1-\frac{\alpha}{2}$ level of probability, m is the number of oils used for calibration, α is the chosen probability level (e.g. $\alpha = 0.05$ gives a 95 % prediction interval), s is the sample standard deviation of the residuals for the oils used for calibration and h is the leverage of the value being considered. Leverage is a measure of how far a value is from the

origin. If the data are mean centred, as in this case, it measures how far a value is from the mean of the calibration data. The leverage shows how much influence the point has on the line of best fit, and is often useful in the identification of outliers(*13*).

The leverage may be calculated via equation A.23(*8*) from the mean-centred predicted fruitiness intensity values.

**Table A.12.** Leverage of predicted fruitiness intensity values calculated for the ten olive oils. Values further from the mean of the data (mean = 5.06) have higher leverage.

| | Predicted value | Leverage |
|---|---|---|
| Oil 1 | 2.66 | 0.127 |
| Oil 2 | 6.34 | 0.0357 |
| Oil 3 | 3.68 | 0.0419 |
| Oil 4 | 8.94 | 0.33 |
| Oil 5 | 4.26 | 0.0142 |
| Oil 6 | 6.97 | 0.0796 |
| Oil 7 | 6.45 | 0.0421 |
| Oil 8 | 4.15 | 0.0183 |
| Oil 9 | 1.38 | 0.298 |
| Oil 10 | 5.81 | 0.0124 |

$$\mathbf{H} = \hat{\mathbf{y}} \, (\hat{\mathbf{y}}^{\mathbf{T}} \, \hat{\mathbf{y}})^{-1} \, \hat{\mathbf{y}}^{\mathbf{T}} \qquad \text{(A.23)}$$

where **H** is the leverage matrix. The leverage value for each point is the corresponding diagonal element in the **H** matrix. All off-diagonal elements are discarded and the leverage value for each oil is stored in the leverage vector **h**. The leverage is included in the prediction interval calculation because values further

from the mean of the calibration values are predicted with less certainty. The leverage is a convenient measure of how far the values are from the mean to ensure the prediction interval is larger for more distant values to reflect the greater uncertainty.

PLS regression uses interpolation, not extrapolation. The VOC concentrations of subsequent oils must be within the ranges of those used for calibration if reliable predictions are to be obtained(*13*). As implied above, the best possible predictions are obtained when all VOC concentrations are equal to the means of the oils in the calibration set.

To predict fruitiness intensity values and calculate prediction intervals for new oils from their VOC concentrations, equations A.20 and A.22 are used. The new VOC concentrations (**X** matrix), standardised concentrations ($\check{\mathbf{Z}}$ matrix), predicted standardised fruitiness values ($\hat{\mathbf{y}}$ vector), predicted fruitiness values ($\hat{\mathbf{y}}_{\mathbf{un}}$ vector) and lower and upper bounds of the prediction intervals for five new olive oils are shown in table A.13. In order to obtain the correct leverages of the new values, the mean-centred predictions of the calibration data ($\hat{\mathbf{y}}_{\mathbf{mc\ cal}}$, the predictions for the original ten oils) must be used alongside those of the new data ($\hat{\mathbf{y}}_{\mathbf{mc\ new}}$) as shown in equation A.24.

**Table A.13.** Relevant input data and predictions of fruitiness values from VOC concentrations for five new olive oils. The **b** vector from table A.8 was used in equation A.20 to obtain the $\hat{\mathbf{y}}$ vector of fruitiness values. All VOC concentrations and subsequent predicted fruitiness values are inside the ranges of their corresponding variables for the ten oils in the calibration set.

| New **X** | | **Ž** | | $\hat{\mathbf{y}}$ | $\hat{\mathbf{y}}_{\mathbf{un}}$ | Prediction Intervals (95 %) | |
|---|---|---|---|---|---|---|---|
| VOC 1 | VOC 2 | VOC 1 | VOC 2 | | | Lower limit | Upper limit |
| 14.1 | 6970 | -0.397 | -0.282 | -0.337 | 4.31 | 3.59 | 5.02 |
| 18.6 | 7650 | 0.246 | 0.13 | 0.187 | 5.48 | 4.78 | 6.18 |
| 12.1 | 6380 | -0.68 | -0.634 | -0.653 | 3.6 | 2.86 | 4.33 |
| 25.9 | 9530 | 1.27 | 1.25 | 1.26 | 7.88 | 7.14 | 8.62 |
| 23.7 | 8800 | 0.958 | 0.816 | 0.882 | 7.04 | 6.32 | 7.75 |

$$H = \hat{\mathbf{y}}_{\mathbf{mc\ new}}\left(\hat{\mathbf{y}}_{\mathbf{mc\ cal}}^{\mathbf{T}}\hat{\mathbf{y}}_{\mathbf{mc\ cal}}\right)^{-\mathbf{1}}\hat{\mathbf{y}}_{\mathbf{mc\ new}}^{\mathbf{T}} \qquad \text{(A.24)}$$

The predicted results may now be plotted in a chart similar to those of the calibration data in figure A.8, although in this case there are no measured fruitiness data for evaluation of the predictions (figure A.9):

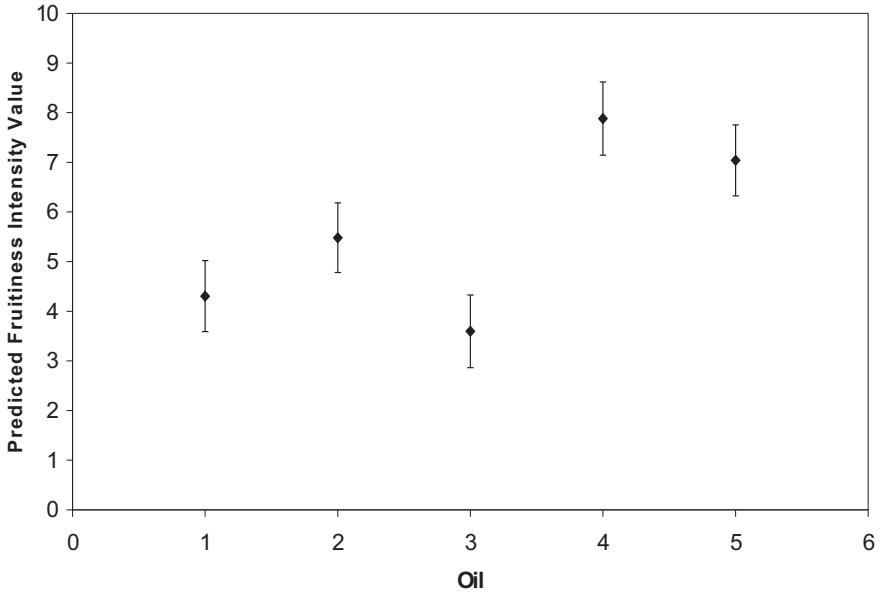



**Figure A.9.** Predicted values of fruitiness intensity plotted with prediction intervals at 95 % confidence from table A.13.

## A.4 References


(1) Reyment, R.; Jöreskog, K. *Applied Factor Analysis in the Natural Sciences*; Cambridge University Press: New York, NY, USA, 1993.

(2) Beebe, K.; Pell, R.; Seasholtz, M. *Chemometrics: A Practical Guide*; John Wiley & Sons: New York, NY, USA, 1998.

(3) Borgognone, M. G.; Bussi, J.; Hough, G. Principal component analysis in sensory analysis: covariance or correlation matrix? *Food Qual. Pref.* **2001**, *12*, 323-326.

(4) Principal Components and Factor Analysis. http://www.statsoft.com/textbook/stfacan.html (accessed Apr 2, 2007)

(5) Franke, J. E. Quantitative Analysis, Inverse Least Squares and Classical Least Squares Methods for Quantitative Vibrational Spectroscopy. In *Handbook of Vibrational Spectroscopy*; J. Chalmers and P. Griffiths, Eds.; John Wiley & Sons Ltd.: New York, NY, USA, 2002.

(6) Harman, H. H. *Modern Factor Analysis*; The University of Chicago Press: Chicago, IL, USA, 1976.

(7) Haaland, D.; Thomas, E. Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Anal. Chem.* **1988**, *60*, 1193-1202.

(8) Gemperline, P. *Practical Guide to Chemometrics*, 2nd ed.; Taylor & Francis: Boca Raton, FL, USA, 2006.

(9) Höskuldsson, A. Dimension of linear models. *Chemom. Int. Lab. Sys.* **1996**, *32*, 37-55.

(10) Wise, B.; Gallagher, N.; Bro, R.; Shaver, J.; Windig, W.; Koch, R. *PLS_Toolbox Version 3.5 for use with MATLAB™*; Eigenvector Research, Inc.: Manson, WA, USA, 2005.

(11) Martens, H. A.; Dardenne, P. Validation and verification of regression in small data sets. *Chemom. Int. Lab. Sys.* **1998**, *44*, 99-121.

(12) Faber, N.; Song, X.-H.; Hopke, P. Sample-specific standard error of prediction for partial least squares regression. *Trends Anal. Chem.* **2003**, *22*, 330-334.

(13) Kramer, R. *Chemometric Techniques for Quantitative Analysis*; Marcel Dekker, Inc.: New York, NY, USA, 1998.

# Appendix B

## Statistics

### B.1 Standard Deviation

As calculated in Microsoft Excel (Microsoft Corporation, Redmond, WA, USA), the standard deviation of a population as estimated from a random sample from that population (the sample standard deviation, σ) is (equation B.1):

$$\sigma = \sqrt{\frac{\Sigma(x - \text{mean}(x))^2}{n - 1}} \qquad \text{(B.1)}$$

Where the x values are the data from which the standard deviation is to be determined and n is the number of elements in the sample. The sample standard deviation is used extensively throughout this thesis to give an indication of the uncertainty associated with results. From this the Coefficient of Variation (CV) is often derived. The CV is calculated as in equation B.2:

$$CV = 100\,\frac{\sigma}{\text{mean}(x)} \qquad \text{(B.2)}$$

The CV is sometime referred to as the percent standard deviation.

## B.2 Robust Standard Deviation

The International Olive Oil Council defines robust standard deviation ($\sigma^*$) as in equation B.3(*1*):

$$\sigma^* = \frac{1.25\ IQR}{1.35\ \sqrt{n}} \qquad (B.3)$$

Where IQR is the interquartile range (the 75$^{th}$ percentile minus the 25$^{th}$ percentile, or the range spanned by 50 % of the data). The median of sensory panel results is normally reported as opposed to the mean. Sensory panel results are reported here as median ± robust standard deviation.

## B.3 Standard Error of Slope

According to Ott and Mendenhall(*2*), the standard error of the slope of a regression line ($\sigma_m$ ) for a collection of x and y values is given by equation B.4:

$$\sigma_m = \frac{\sigma_\varepsilon}{\sqrt{S_{xx}}} \qquad (B.4)$$

Where $\sigma_\varepsilon$ and $S_{xx}$ are defined in equations B.5 and B.6.

$$\sigma_\varepsilon = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}} \qquad (B.5)$$

Where ŷ is the value of y predicted from the linear regression function (consisting of slope m and intercept c).

$$S_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n} \qquad \text{(B.6)}$$

## B.4 Coefficient of Determination

The coefficient of determination is the square of the Pearson linear correlation coefficient (R), and measures the correlation between two sets of data. The value of the Pearson correlation coefficient is often referred to as simply 'correlation', while its square, the coefficient of determination, is often called the '$R^2$ value'. Correlations may take values between -1 and 1, with 0 indicating no correlation and -1 or 1 indicating that the two sets of data are perfectly proportional to each other(*3*). A correlation of -1 specifies a negative proportionality constant. The $R^2$ value shows how much of the variation seen in one set of data can be explained (or predicted) by considering only the other set. Values of 0 and 1 for $R^2$ have the same meanings as for correlations, however intermediate $R^2$ values are more easily understood as they vary linearly with the proportion of variation shared by the two sets of data(*3*). The $R^2$ value is calculated as in equation B.7(*3*), and a Matlab script which uses this equation is provided in section C.2.6. Microsoft Excel also calculates an $R^2$ value, with the RSQ worksheet function.

$$R^2 = \frac{\left(n(\Sigma xy) - (\Sigma x)(\Sigma y)\right)^2}{\left(n(\Sigma x^2) - (\Sigma x)^2\right)\left(n(\Sigma y^2) - (\Sigma y)^2\right)} \qquad \text{(B.7)}$$

## B.5 References

(1) International Olive Oil Council COI/T.20/Doc. no. 15/Rev. 1 Sensory Analysis of Olive Oil Method: Organoleptic Assessment of Virgin Olive Oil. http://www.internationaloliveoil.org/downloads/orga6.pdf (accessed Feb 13, 2007)
(2) Ott, L.; Mendenhall, W. *Understanding Statistics*, 5th ed.; PWS-Kent Publishing Company: Boston, MA, USA, 1990.
(3) Chase, W.; Bown, F. *General Statistics*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1992.

# Appendix C

## Matlab Programs and Sensory Assessment Sheets

### C.1 Introduction

This appendix contains Matlab programs written for use during the present research. Matlab provided a level of flexibility and transparency which was not possible with other available mathematical programs. As shown in this appendix, Matlab was used to perform Partial Least-Squares regression (PLSfunc.m, Leaveout.m, PLScal.m, Predint.m), Principal Component Analysis (PCA.m) and find lines of best fit for non-linear curves (VI50finder.m, peroxlog.m). Extensive comments were written to enhance usability. Comment is marked by a percent symbol (%) in Matlab.

### C.2 Matlab Programs

#### C.2.1 PLSfunc

```
function [RMSECV,Rsq,factors,W,T,V,P,ypred] = PLSfunc(X,Y)
% function [RMSECV,Rsq,factors,W,T,V,P,ypred] = PLSfunc(X,Y)
% This program carries out PLS regression calibration.
% X and Y are the dependent (VOC concentrations
% where columns are VOCs and rows are oils) and the
% independent (column vector of oil sensory values)
% data matrix and vector respectively. All the useful
% matrices are given along with how many latent variables
% to use going by the advice in the Matlab PLS Toolbox
% manual (stop when RMSECV stops decreasing by at
% least 2%), and by the maximum Rsq (R-squared) value
% for prediction.
%
% This function calls the functions Leaveout
% and PLScal
%
% Neither X nor Y should be standardised, the program
% standardises both itself.
```

```
RMSECV = zeros(size(X,2),1);
PRESS = zeros(size(X,2),1);
ypred = zeros(size(Y,1),1);
h = 1; % h is the number of latent variables being used in the cross-validation

for h = 1:size(X,2)
        [RMSECV(h,:),PRESS(h,:),ypred(:,h),Rsq(h,:)] = Leaveout(X,Y,h);
                % leave-one-out cross-validation for PLS
                % the rest of the program is deciding how many latent variables
          to use
end

RMSEfac = 1; % the variable 'RMSEfac' is the number of latent variables
        % recommended by the RMSECV value
s = 2; % there must be a starting value of s, and it must be greater than 1

while s > 1.02 % this is the condition for limiting the number of latent variables
        RMSEfac = RMSEfac+1; % this is the increment that makes the loop
        different % each time
        if RMSEfac == size(X,2)+1
                'Warning! Warning, Will Robinson!'
                'This model does not converge!' % this has to stand out because
                % non-convergence is a big deal
                break % this quits the while loop if convergence hasn't been
        achieved
        end
        s = RMSECV(RMSEfac-1,:)/RMSECV(RMSEfac,:); % here s tests the
        % hypothesis that another latent variable will lower the error
end

Rfac = find(Rsq==max(Rsq));
RMSEfac = RMSEfac-1; % the last latent variable to be tested is always a bad
idea,
        % so this gets rid of it
factors = [RMSEfac Rfac]; % factors is a row vector of the suggested number of
        % latent variables from each method
h = max(factors); % the maximum recommended number of latent variables is
        % calculated to be on the safe side

Ex = zeros(size(X,1),size(X,2)); % this prepares Ex for filling with standardised
Xlo
        % values

c = 1;

for c = 1:size(X,2); % this part standardises X
        Ex(:,c) = (X(:,c)-mean(X(:,c)))./std(X(:,c));
end
```

```
Ey = (Y-mean(Y))./std(Y); % Ex and Ey must be subbed in here so the PLS
algorithm
        % doesn't have to be written out twice

[W,T,V,P] = PLScal(Ex,Ey,h); % this spits out the model with h latent variables
ypred = ypred(:,1:h); % this gives all the columns of ypred which might be
useful
        % the correct column to use corresponds to the number of latent
variables
        % selected for inclusion
```

### C.2.2 Leaveout

```
function [RMSECV,PRESS,ypred,Rsq] = Leaveout(X,Y,h)
% function [RMSECV,PRESS,ypred,Rsq] = Leaveout(X,Y,h)
% this does PLS leave-one-out cross-validation
% for a model with h latent variables
% Leaveout is called by PLSfunc and itself calls
% PLScal

ypred = zeros(size(X,1),1); % there needs to be as many predictions and
residuals as
        % there are oils
yresd = zeros(size(X,1),1);
m = 1; % m is the number of the oil that's being left out

for m = 1:size(X,1) % this makes sure there are as many residuals as there are
rows of
                % X
        Xlo = X; % X isn't being changed, so Xlo resets each time the loop
resets
        Ylo = Y;
        Xlo(m,:) = []; % this is the 'take one out' step
        Ylo(m,:) = []; % deleting the 'mth' row
        Ex = zeros(size(Xlo,1),size(Xlo,2)); % this prepares Ex for filling with
                % standardised Xlo values

        c = 1;

        for c = 1:size(Xlo,2); % this part standardises Xlo and Ylo each time
before
                % they go through PLS
                Ex(:,c) = (Xlo(:,c)-mean(Xlo(:,c)))./std(Xlo(:,c));
        end

        Ey = (Ylo-mean(Ylo))./std(Ylo); % Ex and Ey must be subbed in here so
the
                % PLS algorithm doesn't have to be written out twice
```

```
        W = zeros(size(X,2),1); % size(X,2) gives the number of columns of X,
and
                % size(X,1) gives the number of rows
        T = zeros(size(X,1)-1,1);
        V = zeros(1,1);
        P = zeros(size(X,2),1); % for the P(:,n) notation, it's faster if the matrix
already
                % exists

        [W,T,V,P] = PLScal(Ex,Ey,h); % this does PLS with h latent variables

        % now the X and Y cases that are validating the model need to be scaled
        % to the model to give correct residuals

        d = 1;

        for d = 1:size(Xlo,2);
                Xleft(:,d) = (X(m,d)-mean(Xlo(:,d)))./std(Xlo(:,d));
        end

        Yleft = (Y(m,:)-mean(Ylo))./std(Ylo);

        b = W*inv(P'*W)*V'; % once the model with the right number of latent
                % variables is made, the output matrices are used to find the
        regression
                % vector
        ypred(m,:) = Xleft*b; % predicting the appropriate y value for each left
out
                % point
        yresd(m,:) = ypred(m,:)-Yleft; % here's the residual
        ypred(m,:) = ypred(m,:).*std(Ylo)+mean(Ylo); % this rescales ypred to
the Y
                % distribution
                % that assumes that the predicted variables have the same
distribution
                % as Ylo, which I made them have, so it seems reasonable
        yresd(m,:) = yresd(m,:).*std(Ylo); % this rescales yresd to the Y
distribution
                % as well, so that all yresd values are comparable with each
        other
end

Rsq = (Y'*ypred-size(Y,1)*mean(ypred)*mean(Y))^2/((sum(ypred.^2)...
        -size(ypred,1)*mean(ypred)^2)*(sum(Y.^2)-size(Y,1)*mean(Y)^2))
        % 3 consecutive full stops in Matlab continues the code onto the next
line
        % it helps to make it easier to read

PRESS = sum(yresd.^2);
```

```
RMSECV = sqrt(PRESS/size(yresd,1));
```

### C.2.3 PLScal

```
function [W,T,V,P] = PLScal(Ex,Ey,h)
% function [W,T,V,P] = PLScal(Ex,Ey,h)
% Performs Partial Least-Squares
% calibration on an X matrix and a
% Y vector according to Haaland
% & Thomas (using the NIPALS algorithm),
% although they called them A and c
% respectively. The input is Ex and Ey
% to make the calculations neater. They're
% really just X and Y. Output is T (X latent
% variable scores), V (Y latent variable
% loadings) and P (X latent variable
% loadings). The process runs for h iterations,
% so comes up with a model that has h latent variables.
% This program is called by PLSfunc and Leaveout.

n = 1;

for n = 1:h;
        wun = (Ex'*Ey)/(Ey'*Ey);
        W(:,n) = wun/sqrt(wun'*wun);
        T(:,n) = (Ex*W(:,n))/(W(:,n)'*W(:,n));
        V(:,n) = (T(:,n)'*Ey)/(T(:,n)'*T(:,n));
        P(:,n) = (Ex'*T(:,n))/(T(:,n)'*T(:,n));
        Ey = Ey-T(:,n)*V(:,n);
        Ex = Ex-T(:,n)*P(:,n)';
end
```

### C.2.4 Predint

```
function [PredInt] = Predint(Yun,Yest,t,F,LV)
% function [PredInt] = Predint(Yun,Yest,t,F,LV)
% Gives the prediction interval for the
% t value corresponding to the desired level of
% confidence.

% The leverage was calculated from Gemperline,
% Practical Guide to Chemometrics, 2006, while the
%  prediction interval is from Faber, Song and Hopke,
% Sample-specific standard error of prediction for partial
% least squares regression, Trends Anal. Chem.,2003,22,
% 330-334.
```

```
Yestmc = Yest-mean(Yest); % Yest has to be mean-centred so that leverage is
the
        % distance from the mean, not the origin
H = Yestmc*inv(Yestmc'*Yestmc)*Yestmc';
h = diag(H) % h is the vector of leverage values for the Yest values
e = Yun-Yest;
m = size(Yun,1);
sy = std(e);

r = 1;
PredInt = zeros(m,2);

for r = 1:m
        PredInt(r,1) = Yest(r,:)-t*sy*sqrt(1+h(r,:));
        PredInt(r,2) = Yest(r,:)+t*sy*sqrt(1+h(r,:));
end
```

### C.2.5 PCA

```
function [C,eigenvalues,F] = PCA(X)
% function [C,eigenvalues,F] = PCA(X)
% Performs Principal Component Analysis
% using the algorithm presented in Reyment
% and Joreskog. Needs an X matrix called
% X to exist in the workspace, and produces
% L, C and F matrices. These can be
% copied into Excel and formatted using
% Data > Text to Columns. I've flipped the
% output matrices here so the most
% important ones are at the left and the
% largest eigenvalues are at the top of
% the vector.

% G matrices have variables (VOCs or sensory
% attributes) for columns and oils for rows and
% are mean centred. X can be standardised or
% not but it will always be mean-centred by
% this program. S is the covariance matrix,
% U is the eigenvector matrix (principal
% component axes with unit lengths), L is a
% vector of the eigenvalues, C is the PC
% loadings matrix and F is the PC scores matrix.
% Once everything else works out, look at the
% most relevant columns of the F matrix and
% you're done.

G = zeros(size(X));
```

```
[m,n] = size(G);
d = zeros(1,n);
s = 1;
r = 1;

for s = 1:n
        G(:,s) = X(:,s)-mean(X(:,s)); % mean centre X to make G
end

for r = 1:n
        d(:,r) = 1/std(G(:,r)); % make d, the row vector of G variable std devs
end

D = diag(d); % D is the diagonal matrix of inverse G variable std devs
S = G'*G/(m-1);
[U,L] = eig(S);
A = U*sqrt(L);
C = fliplr(D*A);
eigenvalues = flipud(diag(L)); % flip up/down eigenvalues to get the largest at
the top
F = fliplr(G*A*inv(L));

plot(eigenvalues,'^')
xlabel('Eigenvector')
ylabel('Respective Eigenvalue')
title('Scree Plot of Eigenvalues')
```

### C.2.6 Rsq

```
% Calculates an R-squared value for an estimate.
% The values need to be in a column
% vector called Y, and the estimated values
% vector needs to be called Yest. Y and Yest
% must both have the same number of elements.
% If all these conditions are met, the script
% spits out Rsq and everyone's happy.

Rsq = (Y'*Yest-size(Y,1)*mean(Yest)*mean(Y))^2/((sum(Yest.^2)...
        -size(Yest,1)*mean(Yest)^2)*(sum(Y.^2)-size(Y,1)*mean(Y)^2))
```

### C.2.7 VI50finder

```
function [VI50,Rsq,SSE,newvol] = VI50finder(vol,inh)
% [VI50,Rsq,SSE,newvol] = VI50finder(vol,inh)
%
% Finds the VI50 (50% inhibition volume) for a
```

```
% sample from inhibition percentages at different
% volumes. Also gives the volumes as proportions of
% the VI50, an R-squared value for the fit with the
% standard TOSC function of y = 97.6526(1-e(-0.7178x))
% and the sum of squared errors for the fit.
% This function calls leasqr.m (and therefore also dfdp.m)
% and peroxlog.m.
% vol and inh are vectors of oil volumes and percent inhibition
% at those volumes respectively. They may be row or column
% vectors.

% leasqr.m and dfdp.m are available from:
% http://www.koders.com/matlab/fidED579F3A98BB8D09BEB...
% A8743D66060E49797F54A.aspx
% and
% http://www.koders.com/matlab/fid096B87F4F56E86082AB091...
% A8AC0DCE9D38C26029.aspx
% respectively

if size(vol,2) > 1 % if vol and/or inh are row vectors, switch them to column
vectors
        vol = vol';
end

if size(inh,2) > 1
        inh = inh';
end

[Yest Pest] = leasqr(vol,inh,[100,-0.2],'peroxlog'); % find the best fit
        % this doesn't always work, 100 and -0.2 aren't the best guesses for all
data,
        % they might need to be changed sometimes
VI50 = log(1-50/Pest(1,1))/Pest(2,1); % find VI50 from the best fit equation
newvol = vol/VI50; % express the volumes as proportions of VI50

newinh = zeros(size(inh));
k = 1;

for k = 1:size(inh,1)
        newinh(k,1) = 97.6526*(1-exp(-0.7178*newvol(k,1))); % predict the
                % inhibitions from the proportions of VI50 to see how closely
        the
                % function follows the standard function
end

Y = newinh;
Rsq = (Y'*Yest-size(Y,1)*mean(Yest)*mean(Y))^2/((sum(Yest.^2)...
        -size(Yest,1)*mean(Yest)^2)*(sum(Y.^2)-size(Y,1)*mean(Y)^2)); %
find Rsq
SSE = sum((Y-Yest).^2); % and the sum of squared errors
```

### C.2.8 peroxlog

```
function Y = peroxlog(X,P)
% function Y = peroxlog(X,P)
% This is an equation reference program for
% leasqr.m. It is for use on SIFT-MS-TOSC
% assay results in conjunction with VI50finder.m.

Y = P(1)*(1-exp(P(2)*X));
```

## C.3 Olive Oil Sensory Assessment Sheets

### C.3.1 Local Panel Assessment Sheet

The group of local tasters assembled at Syft Technologies assessed olive oil samples using the same scoring system as used for the Mario Solinas olive oil awards run by the IOOC. Olive oils are awarded total scores out of 100 based on the perceived intensities of individual positive attributes. Only extra virgin olive oils are accepted into the Mario Solinas awards, so no undesirable attributes are assessed in this scoring system. The scoring sheet used is shown in table C.1.

**Table C.1.** The scoring sheet used by the local sensory assessors, adapted from the tasting sheet used by the IOOC for the Mario Solinas olive oil awards.

| **Olfactory Sensations/Aroma** | | Oil 1 | Oil 2 | Oil 3 | Oil 4 | Oil 5 |
|---|---|---|---|---|---|---|
| Olive fruitiness | 0-7 | | | | | |
| Other fruits | 0-3 | | | | | |
| Green (grass/leaves) | 0-2 | | | | | |
| Other positive sensations | 0-3 | | | | | |
| Harmony/balance | 0-20 | | | | | |
| Describe aromas (e.g. banana, cut grass, cinnamon, etc.) | | | | | | |
| Partial Score (Max 35): | | | | | | |

| **Gustatory - retronasal sensations** | | | | | | |
|---|---|---|---|---|---|---|
| Olive fruitiness | 0-10 | | | | | |
| Sweet | 0-4 | | | | | |
| Bitter | 0-3 | | | | | |
| Pungency | 0-3 | | | | | |
| Green (grass/leaves) | 0-2 | | | | | |
| Other positive sensations | 0-3 | | | | | |
| Harmony/balance | 0-20 | | | | | |
| Describe flavours | | | | | | |
| Partial Score (Max 45): | | | | | | |

| **Overall** | | | | | | |
|---|---|---|---|---|---|---|
| Complexity | 0-10 | | | | | |
| Persistence | 0-10 | | | | | |
| Partial Score (Max 20): | | | | | | |

| Total Score: | | | | | |
|---|---|---|---|---|---|

86-100 = Gold Medal
76-85 = Silver Medal
65-75 = Bronze Medal

### C.3.2 Olives New Zealand Sensory Panel Assessment of Olive Oil

The Olives New Zealand Sensory Panel based at Hort Research in Auckland, New Zealand assessed olive oil samples using the standard IOOC assessment system for the classification of olive oil. The sheet used is shown in table C.2(*1*). The intensity of each attribute featured on the sheet is rated relative

to known reference standards. On each line the panellist makes a mark a certain distance from the left end of that line. This distance corresponds to the perceived intensity of the attribute in question. The total length of each line at full size is 10 cm, and the intensity of each attribute is taken as the distance (in cm) from the left end of the corresponding line.

Oils are classified by the median of the results for each attribute as judged by between eight and twelve panellists. Each class of oil is defined by certain characteristics, for example extra virgin olive oils must score zero for all defects and more than zero for fruitiness.

**Table C.2.** The scoring sheet used for olive oil classification by all IOOC accredited sensory panels, including the Olives New Zealand sensory panel.

**PROFILE SHEET**
**(for use by tasters)**

**INTENSITY**

**PERCEPTION OF DEFECTS:**

Fusty

Musty

Winey - Vinegary - Acid - Sour

Muddy sediment

Metallic

Rancid

Others (specify)

**PERCEPTION OF POSITIVE ATTRIBUTES:**

Fruity

Bitter

Pungent

## C.4 References

(1) International Olive Oil Council COI/T.20/Doc. no. 15/Rev. 1 Sensory Analysis of Olive Oil Method: Organoleptic Assessment of Virgin Olive Oil. http://www.internationaloliveoil.org/downloads/orga6.pdf (accessed Feb 13, 2007)